José Guilherme Pereira de Almeida Santos

Licenciado em Ciências Biomédicas

Molecular tools to dissect the role of Dmrt2a and Dmrt2b in the left-right axis formation in zebrafish

Dissertação para obtenção do Grau de Mestre em Genética Molecular e Biomedicina

Orientadora: Maria Leonor Tavares Saúde, Professora Doutora,
Instituto de Medicina Molecular
Faculdade de Medicina da Universidade de Lisboa

Júri:

Presidente: Prof. Doutor José Paulo Nunes de Sousa Sampaio

Arguente: Prof. Doutora Solveig Thorsteidóttir

Vogal: Prof. Doutora Maria Leonor Tavares Saúde

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro, 2013

Molecular tools to dissect the role of Dmrt2a and Dmrt2b in the left-right axis formation in zebrafish

Submitted by

José Guilherme Pereira de Almeida Santos

Thesis to obtain the degree of

Masters in Molecular Genetics and Biomedicine

November 2013

Molecular tools to dissect the role of Dmrt2a and Dmrt2b in the left-right axis formation in zebrafish

Acknowledgements

First of all I would like to thank my supervisor Leonor Saúde for taking the risk of giving me the chance to do this work. I would like to thank all your help and for making us believe that everything is going to work fine. I really enjoyed this year in UDEV.

I also would like to thank all my colleagues, Margarida Figueira, “buddy”, João I miss you so much :p, Rita Pinto this project is haunted haha, Ana Ribeiro espectacular, Rita Serrano and Susana Pascoal the forbidden bench, Lara and Sara Matos I promise I will free some space. A lot! Thank you all for helping me during this year.

And now the winners are… HAHA In no particular order, Margarida Pereira, who taught me so much and helped me put this work on the right track, Aida who helped me a lot with fish maintenance, crosses, and sexual dimorphism prediction LOL, and Sara Fernandes, for the wisdom and patience.

I also want to thank Rita Fior and Raquel Mendes for showing me last year how cool developmental biology could be.

Although their effort is not presented in this work, Andreia and Ana Farinho also helped me during this year. Napkins for science!!

Internationally HAHA, I would like to thank Ben Feldman for the amount of extremely helpful emails.

I want to thank my family, Papá, Néné, Tia Susana, Enio and XicoZé, for always supporting me, I want to thank the “Salgados” and “Rodrigues” specially Leo and Duda for treating me so well, and finally I want to thank Dani who always pushes me and makes me move forward.

Abstract

We tend to view the vertebrate body as bilaterally symmetric, but in fact, this only happens from the outside. Internally, most of the organs from heart to liver are asymmetrically positioned. Skeleton and its associated muscles, symmetric structures of the vertebrate body, have its origins in the transient symmetric blocks of mesoderm called somites whereas the asymmetric morphogenesis of the internal organs is due to asymmetric gene expression in the lateral plate mesoderm (LPM).

Previous studies using Morpholino (MO) technology have shown that *dmrt2a* is involved in these two processes in zebrafish. When Dmrt2a levels are reduced, asymmetric gene expression in the LPM becomes randomized and symmetric gene expression in the presomitic mesoderm (PSM) is disrupted. The paralogous of *dmrt2a*, the fish specific *dmrt2b* has been shown to be involved in regulating asymmetric gene expression in the LPM as well.

Here we used the recent Transcription activator-like effector nucleases (TALENs) technology to generate *dmrt2a* and *dmrt2b* mutant alleles that will allow us in the future to uncover the downstream effectors of these transcription factors using *high-throughput experiments.* In addition, we overexpressed *dmrt2a* at the one-cell stage to characterize asymmetry versus symmetry phenotypes.

The results show clearly the ability of TALEN technology to generate mutant alleles in zebrafish. Nevertheless, *dmrt2a* and *dmrt2b* homozygous mutants developed so far fail to recapitulate their previously described MO phenotypes which raise the question on what molecular mechanism(s) allow(s) zebrafish to cope with frameshift mutations.

The overexpression of *dmrt2a* shows that a time window of opportunity during which symmetric embryonic territories are able to respond to asymmetric signals does exist during embryonic development.

Key words: dmrt2a, dmrt2b, left-right, symmetry, asymmetry, TALENs

Resumo

Olhamos geralmente para os vertebrados como sendo simétricos. No entanto, isto apenas é correcto de um ponto de vista exterior, visto que no interior dos vertebrados, a maioria dos órgãos desde o coração ao fígado estão posicionados de forma assimétrica. O esqueleto e os seus músculos associados, estruturas simétricas dos vertebrados, têm a sua origem em estruturas transientes chamadas de sómitos. Já a morfogênese assimétrica dos órgãos internos, deve-se à expressão genética assimétrica na mesoderme da placa lateral.

Estudos anteriores mostraram através do uso de Morpholinos que o gene *dmrt2a* está envolvido nestes dois processos em peixe-zebra. Quando os níveis de Dmrt2a são reduzidos, a expressão genética assimétrica na mesoderme da placa lateral apresenta-se randomizada, enquanto a expressão genética na mesoderme pré-somítica deixa de ser simétrica. O gene *dmrt2a* tem um parálogo no peixe, o *dmrt2b.* Este último também está envolvido na manutenção da expressão genética assimétrica na mesoderme da placa lateral.

Neste trabalho usamos a recente tecnologia TALEN para criar alelos mutantes para os genes *dmrt2a* e *dmrt2b* que nos vão possibilitar no futuro identificar os genes alvo destes factores de transcrição usando metodologias de larga escala.

Além disto, sobre-expressamos o gene *dmrt2a* no estádio de uma célula para caracterizar o fenótipo assimetria vs. simetria.

Os resultados claramente demonstram a capacidade da tecnologia TALEN para criar alelos mutantes em peixe-zebra. No entanto os mutantes homozigóticos desenvolvidos até agora, não revelaram um fenótipo semelhante ao que já havia sido descrito utilizando Morpholinos, o que lança a pergunta sobre qual será, ou quais serão o(s) mecanismo(s) que o peixe-zebra usa para lidar com mutações que alterem a grelha de leitura de um gene.

A sobre-expressão de *dmrt2a* mostrou que a janela de oportunidade durante a qual territórios embrionários simétricos podem responder a sinais assimétricos efectivamente existe durante o desenvolvimento embrionário.

Palavras chave: dmrt2a, dmrt2b, esquerda-direita, simetria, assimetria, TALENs

Contents

List of Figures

List of Tables

Abbreviations

AAD	Acidic activation domain
DSB	Double-stranded breaks
HR	Homologous recombination
HRM	High resolution melting
HRMA	High resolution melting analysis
LPM	Lateral plate mesoderm
MO	Morpholino
NECD	Notch extracellular domain
NHEJ	Non-homologous end joining
NICD	Notch intracellular domain
NLS	Nuclear localization signals
PSM	Presomitic mesoderm
RA	Retinoic acid
RVD	Repeat-variable di-residues
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
TGFβ	Transforming growth factor beta

Chapter 1

Introduction

1.1 Background

It is amazing to realize that the human body is not all about symmetry. We tend to view all the vertebrates as bilaterally symmetric, but that only happens from the outside. Internally, most of the organs from heart to liver are asymmetrically positioned. It is a feature that is conserved throughout chordate evolution, and although varying among different species, the normal individuals within a given species show the same kind of asymmetries. Among higher mammals, left-right asymmetry extends even higher to the brain and nervous system (Levin 2004).

The external bilateral symmetry of the vertebrate body, resides on the skeleton and its associated muscles. These, have their origins in the somites which are transient embryonic structures, formed in pairs along the anterior-posterior axis and in a cyclic and symmetric way. Upon formation, somites then differentiate, giving rise to axial skeleton and skeletal muscles.

The origins of the internal asymmetry can be traced back to the gastrulation stage, before any morphological asymmetries can be observed. During this stage, asymmetric gene expression in the node is initiated, and a conserved cascade of asymmetrically expressed genes referred to as the *nodal-lefty-pitx2* cassette will lead to the morphological asymmetric organization of the organs.

1.2 Left-right asymmetry within the vertebrate body

With an incidence of 1 over 8 000 live births, there are 3 basic types of human laterality problems that can occur: *situs inversus*, where the internal organs are a complete mirror-image of the normal situation. This happens rarely, almost 1 per 20 000 individuals and since all the organs are reversed, the health of the individual is almost not affected, leading to the possibility that the condition remains under-detected. The same does not happen in isomerism. This condition is characterized by a loss of asymmetry. Either two left sides (left isomerism) or two right sides (right isomerism) are formed. As well as isomerism, single organ inversions, where only one organ is misplaced (dextrocardia), have serious health consequences (Figure 1) (Levin 2004).

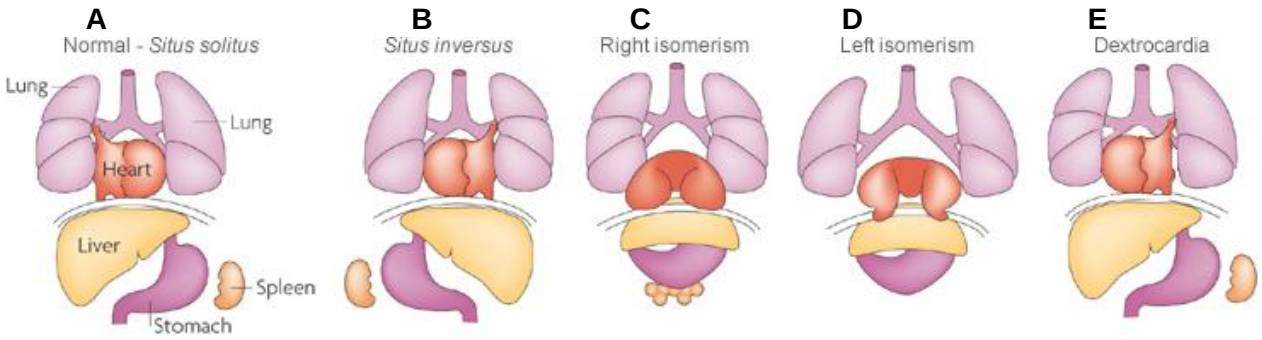


Figure 1.1 Human laterality problems. (A) Normal organization of the internal organs referred to as *situs solitus*. (B) *Situs inversus*. The position of the internal organs is a complete mirror-image of the normal situation. (C) Right isomerism. Two right sides are formed. (D) Left isomerism. Two left sides are formed. (E) Dextrocardia. The heart is the only organ that is reverted from the normal situation. Adapted from (Fliegauf et al. 2007).

It is therefore a challenge to modern scientists to answer a fundamental question: what is the driving force that triggers left-right patterning among so many different species and to what extent it is evolutionary conserved. Also hidden behind this question is the mechanism that ensures that at the same time an internal asymmetric body is being formed, external bilateral symmetry is not compromised.

Being the asymmetrical organization of the internal organs the most visual event, it can only be explained by earlier mechanisms of symmetry breaking. As so, the development of an asymmetric body plan is divided into 3 distinct steps. First, bilateral symmetry has to be broken, forming a left-right axis that is oriented relative to the dorsal-ventral and anterior-posterior axes. Then, differential gene expression between the two sides has to be triggered. And finally, in response to this differential gene expression, changes in cell behavior, such has migration rates, will contribute to a morphological difference between the left and right sides of the body (Vandenberg and Levin 2013).

Differential gene expression between the two sides has one common and conserved feature among different phyla which is the Nodal-Lefty-Pitx2 cassette. At the onset of gastrulation, both around the mouse node, the chicken Hensen's node, the *Xenopus* gastrocoel roof plate and the zebrafish Kupfer's vesicle, Nodal activity becomes restricted to the left side. The expression of *nodal*, a gene encoding a member of the transforming growth factor beta (TGFβ) family, spreads out to the lateral plate mesoderm (LPM), and activates not only the expression of the genes *lefty1*, *lefty2* and *pitx2* but also his own. Lefty2 encodes another member of the TGFβ family, competing for the same Nodal receptors, but unlike Nodal that functions has a dimer, Lefty2 functions has a monomer which allows much more diffusion than Nodal, thus limiting Nodal's activity to the left side. The genetic program leading to subsequent asymmetries is thought to be triggered by Pitx2, a paired-like homeodomain transcription factor that is the effector of Nodal signaling (Nakamura and Hamada 2012, Babu and Roy 2013).

Despite the evolutionary conservation of this Nodal activity on laterality, it remains to be explained the first step of forming an asymmetric body plan: Which event triggers the formation of the left-right axis and how is Nodal localization on the left side of the node initiated?

It was thought that the answer to this question resided in the action of cilia (Nonaka et al. 1998). These are microtubule organelles that extend from the surface of many cells and that are present at the node of mouse, as well as in Kupfer's vesicle of zebrafish. In mouse, a leftward flow is created by the rotation of cilia, at the ventral pole of the embryo. This is called nodal flow and when an artificially nodal flow was generated independently from ciliary motility, laterality was also determined (Nonaka et al. 2002). Two models attempt to explain this mechanism. With cilia being divided into motile and immotile, the "two cilia model" defends that motile cilia generate the leftward flow and that this is sensed by nonmotile mechanosensory cilia. These are present at the periphery of the node and it has been observed that only this type of cilia exhibit polycystin-2, a protein thought to be involved in mechanosensation (McGrath et al. 2003). Another perspective comes from the morphogen gradient model which predicts that a hypothetical morphogen is carried out through the leftward flow. This transport has been described in mouse. Called Nodal Vesicular Parcels, these are membrane-sheathed vesicles budding from node cells. Upon release of their content Sonic Hedgehog and retinoic acid could be asymmetrically concentrated. In agreement, both models report an asymmetrical Ca2+ release that would be defining for the following establishment of the Nodal-Lefty-Pitx2 cassette (Fliegauf et al. 2007, Speder et al. 2007).

Cilia as the answer to laterality could also be supported from clinical data. Individuals with ciliopathies such as Kartagener's syndrome, apart from having *situs inversus*, usually suffer from respiratory dysfunctions like chronic rhinosinusitis and bronchiectasis. Cilia are also present in the respiratory epithelium being part of a mechanism called Mucociliary clearance that enables the airways to protect the lungs from harmful substances in the surrounding environment, (Morillas et al. 2007, Babu and Roy 2013).

Nevertheless, cilia flow cannot be the only crucial mechanism defining laterality. Many phyla establish a left-right axis without the help of cilia, including vertebrates like chick and vertebrate mammals like pig. Chick and pig have morphologically asymmetric nodes, consequence of leftward cell movements, with pig not even showing to have cilia in the notochordal plate as well as space for the flow to be generated. Moreover, both in chick and in pig, asymmetrical gene expression domains form hours before the asymmetric gene expression of *nodal* on the left side of the node.(Vandenberg and Levin 2010) Other examples come from zebrafish, one of the animal models where cilia flow in the Kupfer's vesicle is thought to possibly play a key role in left-right patterning. Two distinct studies in zebrafish with mutants of the gene *seahorse*, a gene that is involved in multiple cilia-mediated processes (Kishimoto et al. 2008), report incoherent results. In one case a *seahorse* zebrafish mutant shows very little laterality defects even though its ciliary flow is almost absent. On the other hand another study reveals a seahorse zebrafish mutant with normal cilia, yet, half of the studied fish develop laterality defects (Vandenberg and Levin 2010). Overall there are many studies today that show little support on a causal link between cilia and laterality in many different phyla, and even if a question of conservation is raised, hypothesizing that mouse has a particular role for cilia in establishing laterality, how to explain that mutations for the gene *tmem216*, which encodes a transmembrane protein that is localized in the base of primary cilia, result in

ciliopathies (Valente et al. 2010) but do not seem to affect laterality, both in zebrafish and mice (Vandenberg and Levin 2010).

Results from chick, pig, and other model organisms, raise another interesting question. Could laterality be a case of convergent evolution? It is difficult to accept that a characteristic spanning so many different phyla might have had different origins. Indeed, the quest for the earliest possible event that could break bilateral symmetry, and that could be a common ancestor of laterality, has retrieved new models on its origins.

The ion flux model resides on the idea that an asymmetric distribution of K+ channels and H+ pumps could be the driven by cell chirality during the first embryonic cleavages. This would lead to an accumulation of serotonin on the right side of the embryo where it represses the expression of *nodal* or its homologues, considering the species. Cilia model supporters considered that the studies involved in this model were inconsistent because they were altering pathways that could consequently affect cilia parameters. Nevertheless it was already confirmed a role for serotonin in the early cell cleavages without affecting node precursor cells (Vandenberg and Levin 2013).

Another model focusing on the earliest possible definition of a left-right axis is the chromatid segregation model. During the first cell cleavage the chromatids would be differentially imprinted and segregated. It has been shown in yeast that differentially segregated chromatin, mRNAs and proteins are sufficient to maintain asymmetry. Some similar mechanisms have been found in eukaryotic cells and embryos (Vandenberg and Levin 2013).

Finally the PCP model, based on a highly conserved mechanism used to correctly orient cell division, like in the Drosophila eyes and wings, mammalian kidney and vertebrate limbs, has also been shown to be associated with left-right patterning. First adopted by the cilia model supporters, since the PCP pathway has the ability to properly position cilia in the node, it has been shown that disrupting the PCP pathway in chick causes laterality defects even though chick do not use cilia to define their left-right axis. The same was shown in frog embryos, with the disruption of the PCP pathway on cells that do not contribute to the node causing laterality defects (Segalen et al. 2010, Vandenberg and Levin 2013).

Non-mutually exclusive, these models confirm that the definition of a left-right axis might happen very early in development. Also they leave open the possibility that not just one mechanism is the ultimate responsible for the correct placement of the asymmetrical body components. Might be that different amplification steps of an initial left-right axis definition occur throughout development, not only in a sequential order but also in a redundant manner. This would explain why most of the experiments trying to address laterality, only achieve randomization phenotypes, with only one small percentage of the studied individuals showing complete reversal of its left-right axis (Vandenberg and Levin 2013).

These recent findings that symmetry can be broken very early in development give a special interest to the question of how to maintain a bilateral symmetric body plan at the same time an asymmetric one is being formed. It is strikingly amazing that if the left-right axis is already defined at the first cell cleavages, it is still possible to create a body structure that when viewed from the outside is symmetrical.

1.3 Bilateral symmetry of the vertebrate body

The external symmetrical appearance of the vertebrate body plan is mostly due to the symmetric organization of the skeleton and its muscles. These have its origins in the somites, transient embryonic structures, composed of blocks of epithelial mesoderm, that form in pairs on both sides of the axial structures, neural tube and notochord, and at a time rate and number that is species specific. In zebrafish each new pair of somites is formed in a period of 30 minutes, whereas in chicken this takes 90 minutes, in mouse 120 minutes and in humans this period lasts approximately 4 to 5 hours. (Dequeant and Pourquie 2008, Lourenço and Saúde 2010).

Extensively reviewed, the process behind the formation of the somites is explained by the clock and wavefront model. A set of genes, called the cyclic genes, are continuously oscillating their expression along the presomitic mesoderm (PSM). This was first discovered in chicken with the gene *hairy1* showing a dynamic and cyclic expression pattern along the PSM with the same periodicity of somite formation (Palmeirim et al. 1997). Other cyclic genes were then discovered. These are called *hes* genes in the mouse and *her* in zebrafish and show a similar oscillating pattern of expression along the PSM. The expression of the cyclic genes starts in the most posterior part of the PSM and through what can be observed as 3 distinct phases of expression, it reaches the most anterior part of the PSM, setting the time for the formation of a new pair of somites. At the same time the cyclic genes are oscillating, opposing gradients of Fgf8/Wnt and Retinoic Acid (RA) mark the position of what is called the determination front. Immature cells supplied from the tail bud, posteriorly, are under the influence of Fgf8 and Wnt signaling and they will only start to differentiate when they reach the most anterior part of the PSM. The process is continuously repeated with the determination front moving posteriorly, until all the somites are formed (Figure 1.2) (Dequeant and Pourquie 2008, Lourenço and Saúde 2010).

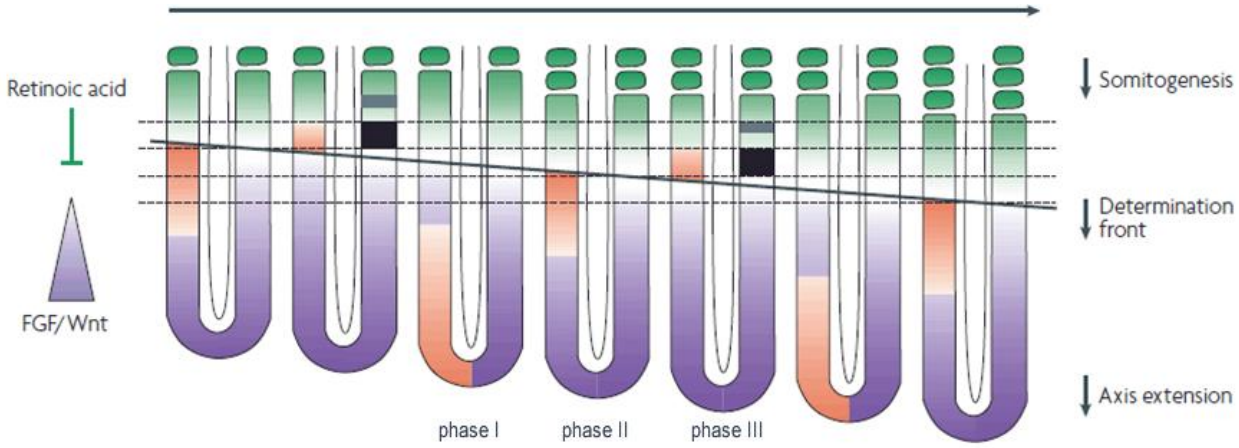


Figure 1.2 Clock and Wave front model of somite formation. The synchronized mRNA oscillations of cyclic genes along the presomitic mesoderm (PSM) describe a wave of expression that is initiated in the posterior region of the PSM (phase I) and moves towards the anterior region of the PSM (phase III) where it slows down culminating with somite formation. At the same time, the determination front marks the position where each new pair of somites is formed. The determination front is defined by opposing gradients of FGF/Wnt and Retinoic acid and moves posteriorly until all the somites are formed. Adapted from (Dequeant and Pourquie 2008).

The synchronized oscillations of the cyclic genes reside mainly on the Notch signaling pathway. The function of this pathway is to coordinate gene expression in contiguous cells. A signal-sending cell expresses a Notch ligand, as it is Delta, and the binding of this ligand with the receptor, Notch, leads to the cleavage of Notch, releasing the Notch intracellular domain (NICD), that translocates to the nucleus where it acts as a transcriptional regulator. The detached extracellular fragment of Notch (NECD) along with Delta is endocytosed into the Delta-expressing cell (Lewis et al. 2009).

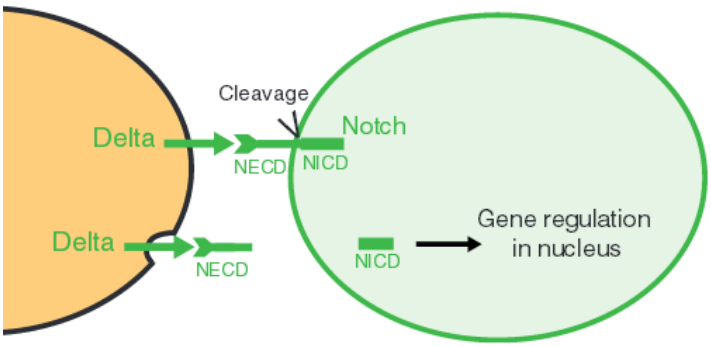


Figure 1.3 Simple representation of cell-cell communication via the Notch signaling pathway. A signal sending cell, expresses the Notch ligand Delta. Once Delta binds the Notch receptor, the Notch intracellular domain (NICD) is released upon cleavage of Notch, and translocates into the nucleus where it acts as a transcriptional regulator. The detached fragment of Notch (NECD) is endocytosed along with Delta into the signal sending cell. Adapted from (Lewis et al. 2009).

The Notch signaling pathway was first associated with the synchronized oscillations of the cyclic genes based on the fact that the *her* genes are targets of the Notch pathway and also by its salt-and-pepper expression pattern in the PSM of notch mutants which could be due to a failure of synchrony in its oscillations. Her1 and Her7, transcription factors, establish a negative feedback loop that leads to a periodic repression of DeltaC. This way, it is possible for neighbouring cells to be synchronized, at the same time they oscillate along the PSM (Holley et al. 2002, Dequeant and Pourquie 2008, Lewis et al. 2009). The cell-cell communication that is the center of this Notch signaling based synchronization was further confirmed with the implantation of cells overexpressing *deltaC* in a zebrafish embryo resulting in the desynchronization of the cyclic gene expression (Ishimatsu et al. 2007).

The role of Notch signaling pathway as a coordinator of this oscillatory gene expression and not so much as an initiator of these oscillations was further evidenced. When zebrafish embryos were treated with the inhibitor DAPT, which inhibits the enzyme that releases the Notch intracellular domain (NICD), blocking signaling transmission, somite defects occurred. Nevertheless this happened with a delay thought to correspond with a gradual disordering of the gene expression pattern (Riedel-Kruse et al. 2007).

Upon formation, the somites undergo a differentiation process. It begins with the formation of different cellular compartments, each one with its unique gene expression, which leads to the development of the different tissue progenitors. Dorsally, the dermomyotome is an intermediary structure that gives rise to the progenitors of the skeletal muscles (myotome), limb muscle progenitors and dermis of the back, whereas ventrally the somite gives rise to the sclerotome, progenitors of the axial skeleton (Hollway et al. 2007).

Zebrafish, supported by its swim bladder, has no use for a robust skeleton. Consequently the zebrafish somite is predominantly composed of myotome, with the sclerotome restricted to a minor fraction of the somite. Although no significant structure resembles the dermomyotome, anterior somitic cells constitute its functional equivalent (Stickney et al. 2000, Hollway et al. 2007).

1.4 Symmetry versus asymmetry during development

At the same time symmetric somites are being formed in the PSM, asymmetric signals are being transferred to the lateral plate mesoderm (LPM). The mechanism that prevents these signals from reaching the PSM and thus disrupting symmetric somite formation resides on RA signaling.

Apart from its role in positioning the determination front during somitogenesis, RA is also involved in this process, buffering these asymmetric signals. Mouse *raldh2* mutant embryos, which lack the enzyme that produces RA, exhibit fewer somites on one side of the PSM (Vermot et al. 2005). This is caused by a desynchronization of the waves of expression along the PSM which leads inevitably to an asymmetric somite formation between both left and right sides. This has also been shown in chick (Vermot and Pourquie 2005) and zebrafish (Kawakami et al. 2005).

In zebrafish, studies using a translation blocking Morpholino (MO), observed that the initiation of somitogenesis is bilaterally symmetric even in the *raldh2* morphants. Nevertheless, as somite formation proceeds, *raldh2* morphants start developing an uneven number of somites between both sides of the PSM. Interestingly, after the 13 somite stage, symmetric somite formation is recovered in the *raldh2* morphants. Consistent with the role of RA to buffer the asymmetric signals from the node, is the fact that in chick, the left-right flow induces a transient lateralization in Fgf, Wnt and Notch pathway components (Boettger et al. 1999, Rodriguez-Esteban et al. 2001, Kawakami et al. 2005).

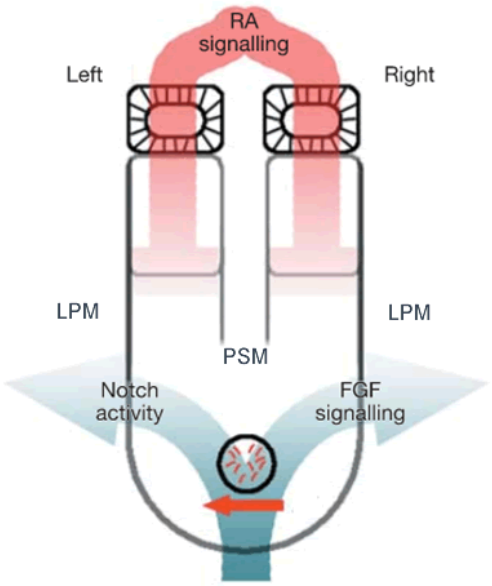


Figure 1.4 Retinoic acid (RA) protects the presomitic mesoderm (PSM) from asymmetric signals. At the same time asymmetrical signals (red arrow) are being transferred from the node to the left lateral plate mesoderm (Left LPM), RA signaling (red) protects the PSM allowing symmetric somite formation. Both Notch and FGF signaling are transiently lateralized (blue), but by the action of RA these regain its symmetric activity in PSM. Adapted from (Kawakami et al. 2005).

To buffer the asymmetric signals, RA also presents a transient asymmetrical signaling. *rere*, encoding a chromatin-remodeling protein, positively regulates RA signaling by forming a complex with the nuclear receptor NR2F2 (COUP-TFII), p300 (EP300) and RARs. *nr2f2*, was found to be asymmetrically expressed in the right PSM. Combining to the fact that a mutation in the mouse *rere* leads to a similar phenotype as the one observed in *raldh2* mutants, a revised model was proposed where the action of RA signaling as a buffer, may itself be transiently asymmetric (Vilhais-Neto et al. 2010).

1.5 The role of *dmrt2a*

The focus of this work, the gene *dmrt2a*, belongs to a family of transcription factors called DMRT (DM related transcription factors). These are genes encoding a zinc finger like DNA binding motif, the DM domain, and although initially associated with sex determination, some of these genes have since been linked with other developmental processes (Hong et al. 2007).

Loss of function experiments, using the MO, as well as gain of function experiments, using mRNA injections, suggested that *dmrt2a* is involved in the correct left-right patterning of the zebrafish embryo. Correct asymmetrical gene expression in the LPM and bilateral symmetric gene expression in the PSM are broken, both in its up regulation and down regulation (Saude et al. 2005, Matsui et al. 2012). The expression pattern of *dmrt2a* correlates with these observations since it is expressed in the Kupfer's vesicle, the laterality organ, and also in the developing somites of zebrafish (Lourenco et al. 2010).

In respect to the LPM, *dmrt2a* knock down or overexpression lead to the randomized expression pattern of genes like *pitx2a* and *spaw*, which normally are restricted to the left side of the LPM. Consequently, the correct positioning of the organs is compromised. Concerning the PSM, the cyclic genes and several other genes like *myoD*, *fgf8*, *raldh2* or *cyp26a*, become bilaterally asymmetric. Nevertheless this asymmetric gene expression is only observed until the 12 somite stage with no distinctive phenotype in later stages of development being observed.

In the mouse, where *dmrt2a* has its homologous gene *dmrt2*, considerably different observations were made. Here, Dmrt2 is not involved in left-right patterning. In the null *dmrt2* mouse *nodal* is restricted to the left side of the node and LPM and *pitx* is restricted to the left side of the LPM. Also, this mutant does not express *dmrt2* in the node which correlates with its normal organization of internal organs (Lourenco et al. 2010). On the other hand axial skeleton and rib patterning defects can be observed in a null *dmrt2* mouse but bilaterally symmetric gene expression of *hes7* is not affected (Seo et al. 2006, Lourenco et al. 2010). The null *dmrt2* mouse has problems in somite differentiation. Here *dmrt2* is specifically expressed in the dermomyotome, and although no significant alterations were detected in terms of muscle development, it is clear that the normal arrangement of the myoblasts in the myotome is affected with the myocytes failing to elongate and occupy the entire rostral-caudal domain of the myotome. These changes may arise from an essential role of *dmrt2* in providing extracellular matrix components within the mouse dermomyotome which are essential for correct myocyte differentiation (Seo et al. 2006).

The gene *dmrt2a* has also a paralogous, the fish specific *dmrt2b*. Similarly to *dmrt2a*, knock down of *dmrt2b* randomizes the asymmetrical gene expression of genes in the LPM leading to an incorrect organ positioning. However this similarity is restricted to the LPM with no involvement of *dmrt2b* in the synchronization of the cyclic genes during somitogenesis. Nevertheless *dmrt2b* is also expressed in the somites and has its role during somite differentiation, since its knock down inhibits the sonic hedgehog pathway leading to slow muscle defects (Liu et al. 2009).

1.6 Transcription activator-like effectors (TALEs)

Transcription activator-like effectors (TALEs) are a class of DNA binding proteins that can be found in some species of plant pathogenic bacteria of the genus Xanthomonas. Consisting on these bacteria key virulence factors, TALEs are translocated into the plant cell cytoplasm via type III secretion system, and have the ability to reprogram host cells by mimicking eukaryotic transcription factors. They enter the nucleus, bind to specific sequences in the host gene promoters and activate transcription of downstream genes (Boch et al. 2009, Cermak et al. 2011).

Structurally TALEs are composed of an N-terminus required for type III secretion, a C-terminus containing nuclear localization signals (NLS) and an acidic activation domain (AAD) common in transcription factors. Their binding specificity comes from a region of typically 33-34 amino acid repeats, followed by a single truncated repeat of 20 amino acids. The number of this repeats may vary but the true singularity of each TALE comes from a repeat polymorphism that exists across these proteins and that is

concentrated at residues 12 and 13, referred to as repeat-variable di-residues (RVD). These are the residues that specify the target, one RVD to one nucleotide, and although many different RVDs can occur across TALEs, four of them, HD, NG, NI, and NN, account for 75% of the total and respectively associate with one of the four DNA bases. Also, and common across TALEs, is the requirement for the binding site to be preceded by a 5'T (Boch et al. 2009, Bogdanove et al. 2010, Cermak et al. 2011).

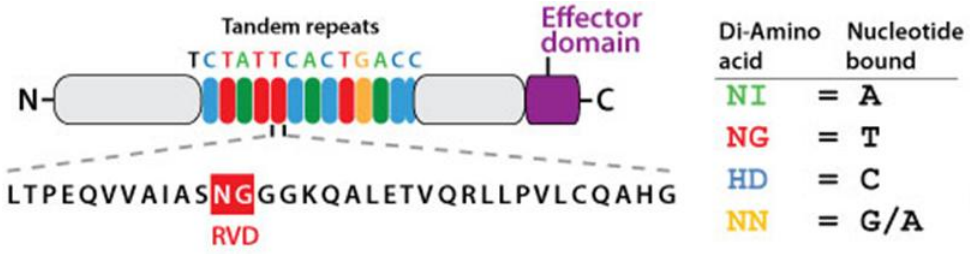


Figure 1.5 Simple representation of Transcription activator-like effectors (TALEs). (A) The N-terminus is required for type III secretion and the C-terminus contains nuclear localization signals (NLS). The effector domain acts as a transcription activator. Binding specificity comes from a region of typically 33-34 amino acid repeats (letters in black). Concentrated at residues 12 and 13 the RVDs (red) specify the nucleotide that is targeted by each repeat. (B) The most common RVDs and theirs target nucleotides.

1.7 Genetic engineering with TALE nucleases (TALENs)

The simple code that governs TALE activity made it possible to customize these proteins in order to achieve modifications in DNA sequences of interest.

Transcription activator-like effector nucleases (TALENs) are TALE-nuclease chimeras, where the TALE region required for high-affinity DNA binding is fused into the catalytic domain of FokI. The result is a gene modification tool capable of inducing double-stranded breaks (DSB) in vivo. It is important to point out that FokI cleaves as a dimer, so two opposite TALENs are needed to create this DSB leaving a spacer in between them so that the two FokI domains can act. (Cermak et al. 2011, Miller et al. 2011) Two major processes can be involved in repairing these DSB: non-homologous end joining (NHEJ), resulting in small insertions or deletions (indels), and homologous recombination (HR), used for sequence modifications, provided by a donor template (Cade et al. 2012).

It is also important to point out that the accuracy of this DSB is in a range of 4 bp. What it means is that the binding sites of TALEN pairs are designed so that the spacer in between them contains the desired cut site, but one cannot be sure about where the exact cut will happen along the spacer sequence. Moreover, from cell to cell, in a given embryo, the cut will happen differently, in a range of 4 bp, giving rise to mosaic embryos in the F0 generation. (Dahlem et al. 2012)

1.8 TALENs assembly

TALENs assembly is the process by which all the RVDs that confer binding specificity to a TALEN are put together in a previously constructed vector that includes, among others, the catalytic FokI domain. Considering that each TALEN monomer has around 16 RVDs and that at least one TALEN pair is needed to produce a cut, it would be a very time consuming, expensive and error-prone protocol to make if traditional molecular cloning techniques were used. This was made easier following the work of (Cermak et al. 2011) where the recent method of Golden Gate cloning was applied to TALENs assembly. Through Golden Gate cloning several separate plasmids can be efficiently cloned into an acceptor vector in one single reaction mixture in one tube. Using type IIS restriction endonucleases which cleave outside their recognition sites (sticky ends), and a ligation enzyme, the reaction mixture is subjected to multiple steps of digestion/ligation, according to the optimal temperatures for each of the enzymes used. Since the correct assembly eliminates the restriction enzyme recognition site, the efficiency of this procedure is high with almost all of the colonies after transformation possessing the desired construct (Engler et al. 2009, Cermak et al. 2011).

1.9 Focus of this work

Previous studies that tried to unveil the function of *dmrt2a* and *dmrt2b* in zebrafish were done using MO (Saude et al. 2005, Liu et al. 2009). Although broadly used in genetic studies MOs have a transient effect and are not transmitted through the germline. Also, MOs may have off-target effects which can lead to phenotype misinterpretation (Huang et al. 2012).

During this work recent TALEN technology (Bogdanove et al. 2010, Cermak et al. 2011, Miller et al. 2011) will be used to try to generate homozygous mutants for *dmrt2a* and *dmrt2b.* With homozygous mutants for these two genes high-throughput analysis using Microarray technology (Sobek et al. 2006) can be made without having the risk of variability which can be introduced by the MO.

Also the possibility of these mutants being viable will allow the generation of a double mutant for *dmrt2a* and *dmrt2b.* This would help us understand better the roles of these two paralogous genes, and possibly unveil to what extent they are related to the homologous mouse *dmrt2.*

Also during this work, an overexpression study of *dmrt2a* will try to assess the existence of a time window of opportunity during which asymmetries are defined.

Chapter 2

Materials and Methods

2.1 Zebrafish lines and maintenance

The zebrafish strain TU was used. Adult fish and embryos were maintained and bred according to standard procedures (Westerfield 2000).

2.1 List of Primers

Primers used during the course of this work, either for cloning or genotyping, are listed in Table 2.1. Except where indicated primers used during this work were designed using NCBI primer blast (NCBI) and oligoanalyzer (Integrated DNA Technologies) and synthesized by Stabvida.

Table 2.1 List of Primers used during this work. Restriction sites highlighted in yellow (EcoRI), blue (StuI) and green (XhoI). Except for *, primers were designed during this work.

Primer name	Primer sequence	PCR (bp)
dmrt2a_FW	5'-ACTCATCGTTTGTTTGACTGCTTT-3'	572 bp
dmrt2a_RV	5'-AGAACCTCTTGTGCCCCTTTAG-3'	
dmrt2b_FW	5'-GAAACATCCAGACTCACAAGCACAGC-3'	421 bp
dmrt2b_RV	5'-CTGCCATCACTCGCTGCCTCTCC-3'	
pCR8_F1 *	5'-TTGATGCCTGGCAGTTCCCT-3'	variable
pCR8_R1 *	5'-CGAACCGAACAGGCTTATGT-3'	
grunwald_FW *	5'-TTGGCGTCGGCAAACAGTGG-3'	variable
grunwald_RV *	5'-ACGTCCCATCGCGTTGCC-3'	
HRM_dmrt2a_FW	5'-GACACGTTACATGCAGGAAAACA-3'	128 bp
HRM_dmrt2a_RV	5'-CTCCACGTCGATCTCAAACTCC-3'	
HRM_dmrt2b_FW	5'-CACAGGTAGATGCGACCCAC-3'	87 bp
HRM_dmrt2b_RV	5'-CTTCATCCGTGCCCATGACC-3'	
HRM_dmrt2a_par2_FW	5'-GCGATGATCAGGCGGTGTTC-3'	74 bp
HRM_dmrt2a_par2_RV	5'-GCGGTCAGATTTGTCGTCGT-3'	
dmrt2a_cloning_FW	5'-TTGGAATTCTATGACGGATCTGTCCGGCAC-3'	1508 bp
dmrt2a_cloning_RV *	5'-AGGCCTTTTTTA CTGAGATTTCCGATTTAAAGAAAGCGC-3'	
dmrt2b_cloning_FW	5'-TTG GAA TTCT ATGTCCACTAAAGCGGATAGGG-3'	1091 bp
dmrt2b_cloning_RV	5'-AATCTCGAGTTTATCTCATGAGCAGTGCCTC-3'	

2.3 Transformation of competent cells

Competent cells, previously prepared in our lab, and kept at -80ºC were thawed on ice, and at the same time, the tube where the bacterial cells would be transformed was cooled. 10μl of a cloning reaction plus 100μl of competent cells were put together and left on ice for 30 minutes. A heat shock was done at 42ºC for 1 minute and then the tube was cooled again on ice for 2 more minutes. 900μl of SOB solution was added and the mixture was then incubated with shaking at 37ºC for 45 minutes. Finally 50μl of the mixture were plated on LB agar media, containing the appropriate antibiotic, and incubated at 37ºC overnight.

2.4 Colony PCR

Throughout this work colony PCR was used to help identify bacterial colonies with the right ligation product prior to miniprepation. 40ml falcons with 5ml of appropriate medium and antibiotic were carefully prepared and identified. Then, PCR tubes containing 10μl of water were also identified with the corresponding designations. On ice, a PCR master mix composed of 12.5μl of Quick-Load Taq 2X Master Mix (New England BioLabs) and 200μM of each primer was also prepared. Using a pippete tip, a colony was picked. The tip was then inserted in a pippete to help mix the colony in the water that was inside the PCR tube by pipetting up and down. The tip was then placed inside the 40ml falcon in contact with the medium. This procedure was repeated for each different colony. Once all the colonies were picked, the PCR master mix was added to each tube and a PCR reaction was started on a thermal cycler according to Quick-Load Taq 2X Master Mix manufacturer's instructions. At the same time, inoculums were incubated with shaking overnight at 37ºC.

Colony PCR products were then run on a 1% Agarose gel (*SeaKem*) and clones possessing the correct ligation product were identified. The corresponding inoculums were left incubating for minipreparation on the next day, whereas the others were discarded. If no correct ligation product was identified, the whole process was repeated with different colonies.

2.5 Cloning of *dmrt2a* and *dmrt2b*

Since both *dmrt2a* and *dmrt2b* are expressed during somitogenesis, total RNA was extracted from 12 hour post fertilization wild type zebrafish embryos (described below) and cDNA synthesized with MMLV-Reverse Transcriptase kit (Promega).

Zebrafish *dmrt2a* and *dmrt2b* full length coding sequences were amplified by PCR, using Phusion High Fidelity Polymerase (Thermoscientific) with the following primer set: *dmrt2a_cloning_FW* and *dmrt2a_cloning_RV*; *dmrt2b_cloning_FW* and *dmrt2b_cloning_RV*. Restriction sites are included in these primers, with both the forward primers having an EcorRI restriction site, *dmrt2a _cloning_RV* having a StuI restriction site and *dmrt2b_cloning_RV* having a XhoI restriction site. Reactions included 10μl 5x Phusion

HF Buffer (Thermoscientific), 2μl of each primer 25μM, 2μl of dNTPS 5μM, 2μl of cDNA, 0,5μl Phusion HF and water up to 50μl. Conditions were 30 seconds at 98ºC, 35 cycles of 10 seconds 98ºc, 30 seconds at 68ºC, 45 seconds at 72ºC, and 10 minutes at 72ºC. Reactions were performed on a thermal cycler (Applied Biosystems 2720).

The amplified fragments were then introduced into separate PCS2+ expression vectors, already available in the lab glycerol stock. First, four double digestions were executed. *dmrt2a* insert and PCS2+ vector were digested separately with EcoRI (New England BioLabs) and StuI (New England BioLabs) restriction enzymes using the appropriate buffer according to the manufacturer's instructions. *dmrt2b* insert and another PCS2+ vector were also digested separately but in this case with EcoRI (New England BioLabs) and XhoI (New England BioLabs) restriction enzymes, and also using the appropriate buffer according to the manufactures. Then, using T4 DNA Ligase (New England BioLabs), ligation of each insert with its corresponding vector was performed. Double digestions of *dmrt2a* insert and its corresponding vector by EcoRI and StuI originated sticky ends (EcoRI) and blunt ends (StuI). Here, the ligation reaction mixture was incubated for two hours at 16ºC. In the case of *dmrt2b* insert and its corresponding vector, double digestions originated only sticky ends (EcoRI and XhoI). The reaction mixture was then incubated also at 16ºC but in this case for only 30 minutes.

Ligation products were used to transform competent cells, as already described. In both cases ampillicin was used as antibiotic. To identify the colonies with the correct ligation product, two distinctive approaches were used. Colonies, presumably containing *dmrt2a* insert were identified by colony PCR, as described before, using the same primers that were used for cloning. In the case of *dmrt2b* colony PCR did not prove to be as effective and so, 10 random colonies were picked and used for minipreparation. Analytical double digestions of those minipreps with EcoRI and XhoI were then performed to remove the *dmrt2b* insert, if present. Digested samples were run on a 1% Agarose Gel and correct clones possessing *dmrt2b* insert were finally identified. Both *dmrt2a* and *dmrt2b* correct clones were then sequenced for confirmation.

2.6 TALENs design and assembly

Prior to TALENs assembly, the RVD sequences were carefully designed. The first step was to genotype wild type zebrafish adults, the ones that laid the eggs used to inject the mRNA. Polymorphisms are common in zebrafish (Bradley et al. 2007), so in order to ensure a correct binding of TALENs, the DNA sequences of the region of interest among various wild type adults had to be carefully analyzed, ultimately leading to a selection of the best possible binding region. Having already chosen a safe region to be targeted, TAL Effector Nucleotide Targeter 2.0 (TALE-NT) was used. At this point, the region of the gene to be targeted was filtered for candidate TALEN pairs, according to the length of the monomers and spacer pretended, and following the rule that every binding site has to be preceded by a 5'T. After choosing a TALEN pair, another tool was used called Paired Target Finder. Here, after entering the

desired TALEN pair, a search through the entire genome, in this case the zebrafish genome, for offside possible targets was performed. (Doyle et al. 2012).

In (Cermak et al. 2011), an assembly protocol as well as a complete set of plasmids containing each of the possible RVDs and backbone vectors to be used was created and deposited in the non-profit repository Addgene. Final vectors specific for zebrafish were also created by (Dahlem et al. 2012) and deposited in the same repository.

Considering the approach used in this work where all the TALENs designed did not exceed 21 RVDs long, the assembly of one full TALEN monomer compromised two steps. During the first step (Golden Gate reaction 1), two independent digestion/ligation reactions were made. The first one accommodated the first 10 RVDS to be used, in an intermediate vector. The second digestion/ligation reaction accommodated the remaining RVDs to be used, minus the last one, in another intermediate vector. After transformation of competent cells using these intermediate vectors, minipreparation of plasmid DNA and sequencing, the second step (Golden Gate reaction 2) joined the RVDs contained in the intermediate vectors, plus the last RVD, into one final backbone vector (Cermak et al. 2011).

The final backbone vectors used in this work, one for the left monomer and one for the right monomer, contained an SP6 promoter and the SV40 polyadenylation sequence. As a result, after the correct assembly of the TALEN construct, mRNA synthesis was achieved through SP6 *in vitro* transcription and the resulting mRNA molecules were securely injected into one cell stage embryos without being degraded (Dahlem et al. 2012).

2.6.1 *dmrt2a* and *dmrt2b* TALENs design

As mentioned above, TALENs were designed using TAL Effector-Nucleotide Targeter (TALE-NT) 2.0 web based tools (Cermak et al. 2011). The R-based tool developed by Jorge Velez (NICHD) was also used, although recently this tool has been considered not needed.

Wild type TU zebrafish adults were genotyped for the genes *dmrt2a* and *dmrt2b*. The amplified and sequenced fragments corresponded to the desired target regions.

The corresponding DNA sequences were then introduced in TALEN Targeter. Custom Spacer/RVD Lengths were used with the spacer ranging from 14 to 17 nucleotides and the RVD lengths ranging from 16 to 21 nucleotides. The G substitute chosen was NN. (It is important to note that recently the NH repeat has been considered to be more effective targeting the G nucleotide). The upstream base chosen to each monomer was T.

The results in the form of txt files were introduced in the R-based tool, which sorted the results according to the rule NG + HD > NI + NN. The resulted txt files from R-based tool were then filtered using Microsoft Excel.

Four TALEN pairs were designed, with two pairs targeting *dmrt2a* and two pairs targeting *dmrt2b*. The objective was to design for both genes, TALEN pairs which spacer would contain the ATG start codon and a TALEN pair that would target a region immediately before the start of DM domain. For the gene

dmrt2a this was successfully achieved whereas for *dmrt2b* no TALEN pairs would respect the rule NG + HD > NI + NN in the ATG region. Two pairs targeting the same region in between the ATG region and the DM domain of the *dmrt2b* gene were then designed.

Finally, Paired Target Finder was used to check if the TALENs designed would somehow have off targets in the zebrafish genome. Here Search a Genome/Promoterome Tab was used. In Pre-loaded Sequence, *Danio rerio* (genome) was chosen and for each TALEN pair, the first monomer was inserted in RVD Sequence 1 and the second monomer was inserted in RVD sequence 2. The Score Cutoff chosen was 3.0 which meant that off targets with a score 3 times higher than the best possible score (our target of interest) would be left out. All the retrieved off targets had scores much higher than the best possible score.

2.6.2 *dmrt2a* and *dmrt2b* TALENs assembly using the Golden Gate cloning system

TALEN constructs were assembled using Golden Gate cloning approach as already summarized. First, the bacterial glycerol stock contained in the Golden Gate TALEN and TAL Effector kit was used for minipreparation of all the DNA plasmids needed. Minipreps were done using QIAprep Spin Miniprep Kit according to the manufacturer's instructions. To ease the following step, every DNA sample was diluted to 150ng/μl, with the minipreps that did not have enough concentration being repeated.

In Golden Gate reaction 1, for each monomer, the first ten RVDs modules were joined together into a pFUS_A vector. The remaining RVDs modules, minus the last one, were joined together into a pFUS_B# vector, with the pFUS_B number corresponding to the amount of RVDs that it would accommodate. Each reaction was composed of 150ng of each module vector plus 150ng of pFUS vector, 1μl of BSA-HF (New England Biolabs), 1μl of T4 DNA Ligase, 2μl of 10x T4 DNA Ligase buffer (New England Biolabs) and water to a 20μl final volume. Reactions were performed on a thermal cycler (Applied Biosystems 2720) with the following conditions: 10x (37ºC/5min + 16ºC/10min) + 50ºC/5min + 80C/5min. Plasmid safe nuclease treatment was then performed, using per reaction, 1μl of 10mM ATP and 1μl of Plasmid safe nuclease (Epicenter), in an incubation period of one hour at 37ºC.

Then, transformation of competent cells was done as previously described, using the appropriate antibiotic, in this case spectinomycin and with the exception that petri plates were supplemented with 40μl of X-Gal (20mg/ml) and 40μl of IPTG (0.8M) to perform a blue/white screening.

The next day, 3 white colonies per plate were picked and PCR colony (previously described) was performed using primers pCR8_F1 and pCR8_R1 to evaluate the correct ligation product. Considering PCR colony results, minipreparation of the corresponding inoculums was made and before moving to the next step, every sample was sequenced and then again, diluted to 150ng/μl. To help predict each desired nucleotide sequence is useful the usage of TAL Plasmids Sequence Assembly Tool (Fine).

In Golden Gate reaction 2, corresponding pFUS_A and pFUS_B#, plus the last RVD containing vectors pLR were assembled into final backbone vectors. These were pCS2TAL3-RR to create right

monomers and pCS2TAL3-DD to create left monomers (Dahlem et al. 2012). Reactions were composed of 150ng of each intermediate vector, 75ng of either pCS2TAL3-RR or pCS2TAL3-DD, 1μl of Esp3I (Thermoscientific) restriction enzyme, 1μl of T4 DNA Ligase, 2μl of 10xT4 DNA Ligase buffer and water to a final volume of 20μl. Reactions were again performed on a thermal cycler with the following conditions: 10x (37ºC/5min + 16ºC/10min) + 37ºC/15min + 80C/5min. Note that in this last reaction, plasmid safe treatment as well as Esp3I denaturation were not needed since the final backbone vector had no homology with the inserted repeats.

The ligation products were again used to transform competent cells. Here, ampicillin plates were used, and once more, X-Gal and IPTG were used to perform a blue/white screening. In the next day, PCR colony was repeated in the same fashion as described before, only this time, with primers pGrunwald_FW and pGrunwald_RV. The correct ligations were assessed and the corresponding inoculums were used for minipreparation. Every sample was again sequenced to confirm the correct assembly of all 8 constructs.

2.7 Genotyping

Adult fish were anesthetized using 1xTricaine solution (MS-222) and caudal fins were cut to extract genomic DNA (described below). After this procedure the fish were kept individually isolated until the results from DNA sequencing or HRM were obtained. Genotyping of zebrafish embryos was also performed. In this case, embryos were previously sacrificed using 25xTricaine solution, and separated individually in different tubes, so that single embryo genomic DNA could be extracted (described below). Here sequencing and HRM results were used to statistically predict the abundance of different alleles in siblings or just to assess the efficiency of TALENs activity. It is important to note that during this work, genomic DNA was never extracted from pools of embryos in order to assess TALENs efficiency or germline transmission since single embryo approach is undoubtedly more informative both quantitatively and qualitatively.

2.7.1 Phenol-chloroform genomic DNA extraction

To extract genomic DNA, a caudal fin from an adult zebrafish or a whole zebrafish embryo was incubated in a digestion buffer [NaCl 5; EDTA 0,5M, TrisHCl 1M pH8; SDS 20%] plus 1μl of Proteinase K (20ng/ul) at 55ºC, 400 rpm. overnight. The next day, the tube was vortexed to homogenize the mixture and left incubating for two more hours. Then 200μl of phenol solution (Sigma) was added and the mixture was again vortexed prior to a 10 minute centrifugation at room temperature. The supernatant was carefully transferred to a new tube and another 200μl of phenol solution was added to repeat the last step. After the second centrifugation and supernatant transfer, instead of phenol solution, phenol-chloroform solution (Sigma) was added, and the mixture was once again homogenized, centrifuged and separated. To precipitate the DNA, 500μl of absolute ethanol was added, with the mixture being homogenized simply by inverting the tube a couple of times, and then kept at -80ºC for 30 minutes. The tube is then centrifuged

at 14 000 rpm during 30 minutes at 4ºC (eppendorf 5430 R) and after, the absolute ethanol is exchanged with previously cooled 70% ethanol, releasing the pellet from the tube wall and then centrifuged again for 10 more minutes. Finally, the pellet was air dried and ressuspended in 50μl of water.

2.7.2 NaOH genomic DNA extraction

An adult zebrafish caudal fin or a whole two day post fertilization zebrafish embryo was incubated with NaOH 50mM for 20 minutes at 95ºC. After incubation, the sample was cooled down on ice for about 2 minutes and TrisHCL 1M pH8 was added. For adult zebrafish caudal fins, 100μl of NaOH and 10μl of TrisHCL were used whereas for two days post fertilization embryos the amounts used were 50μl of NaOH and 5μl of TrisHCL. After adding TrisHCL the mixture was homogenized and centrifuged at room temperature for 5 minutes at 13 000g (eppendorf 5424) and the supernatant was separated and stored. The product was directly used for both PCR and HRM.

2.8 Total RNA extraction

RNA was extracted using TRIZOL reagent (Invitrogen). First, approximately 50 embryos at the developmental stage of interest were dechorionated and frozen in 400μl of TRIZOL at -80ºC. The next day, after thawed, the mixture was homogenized by pipetting it up and down and left incubating at room temperature for 5 minutes. Then 120μl of chloroform (PRONALAB) was added and this time homogenization was achieved by shaking the tube around 20 seconds, being careful not to let it open, and incubating it for 3 minutes at room temperature. The mixture was then centrifuged at 6 000g, for 30 minutes at 4ºC. After centrifugation, the aqueous phase was transferred to a different tube and 300μl of Isopropyl Alcohol was added to it. After incubation for 10 minutes at room temperature this new tube was centrifuged at 10 000g for 15 minutes at 4ºC, precipitating the RNA. Supernatant was removed and the RNA pellet was washed with 600μl of 75% EtOH. The tube was centrifuged again at 10 000g for 15 minutes at 4ºC. Then EtOH was removed and the RNA was dried on ice and ressuspended in water, according to the amount of RNA pellet.

2.9 mRNA synthesis

During the course of this work several different mRNA molecules had to be synthesized for microinjection. Since all of them were cloned into PCS2 or PCS2 derived plasmids, the same strategy was used. In order to produce DNA templates, constructs were linearized with NotI (Fermentas), downstream of SV40 PA terminator. Samples were run on a 1% Agarose gel to separate and extract the desired product. After purification, templates were transcribed with SP6 mMessage mMachine kit (Ambion), according to the manufacturers' instructions. Samples were them purified using Illustra Microspin G-25 Columns (GE Healthcare) and stored at -20ºC.

2.10 Anti-sense mRNA probes synthesis

Anti-sense mRNA probes were used during this work for whole mount *in situ* hybridization. Such probes were synthesized from plasmid templates already available in the lab, either in the form of Miniprep or bacterial glycerol stock. Only the probes for *dmrt2a* and *dmrt2b* were synthesized from plasmid templates cloned during this work.

In order to prepare the DNA templates, each plasmid was linearized in a 50μl reaction mixture including the appropriate 1X buffer and restriction enzyme, 5μg of DNA template, 1X BSA if needed, and water. Reaction mixtures were incubated at 37ºC for 2 hours. Then, 2μl undigested plasmid, 2μl of the digestion mixture, and the full content of the digestion mixture were run on a 1% agarose gel for approximately 1 hour to confirm the linearization of the DNA plasmid and allow for complete dissociation of undigested fragments. Only the band corresponding to the linearized fragment was extracted in a dark room using UV light and purified with cleanup *Wizard*® SV Gel and PCR Clean-Up System.

Anti-sense transcripts were then produced using at best 1μg of purified DNA template, in a 25μl reaction containing 1μl buffer (Roche), 7μl of DTT (Promega), 1xBSA if needed, 2.5μl of DIG (Roche) nucleotides, 1μl of RNasin (Promega), and T7 (Roche). Water was also used only if needed. The mixture was incubated at 37ºC for 3 hours.

Precipitation and purification of the transcripts were done by adding to each reaction mixture 20.5μl of water, 2ul of EDTA 0.5M, 2.5μl of LiCL 8M, 150ul of absolute ethanol with an incubation at -20ºC overnight. The next day, a centrifugation at 4ºC for 25 minutes was performed and after removing the excess, 150μl of 70% ethanol was added with an extra 10 minute centrifugation at 4ºC. After removing once again the excess, pellets were dried on ice for 20 minutes and 30μl of EDTA 10mM were used to resuspend the pellet. Anti-sense probes were then stored at -20ºC.

2.11 Whole mount *in situ* hybridization

Zebrafish embryos were collected, according to the developmental stage to be assessed, and fixed in 4% paraformaldehyde solution prepared in PBS, overnight at 4ºC or 3 hours at room temperature. In order to be stored the embryos were dehydrated. First, two washes with 0,1% Tween 20 (sigma) in PBS (PTW) were made, and then the embryos were transferred sequentially to 50% MetOH (sigma) solution in PTW and to 100% MetOH. These embryos were then kept at -20ºC for two hours before being used, or stored, also at -20ºC.

In a protocol that is divided in three days, the first day started with rehydration. Here, the embryos were washed with 75%, 50%, and 25% MetOH solutions in PTW, and then with 100% PTW for four times, with each wash lasting for 5 minutes. Chorions were then removed, using 1% agarose in embryo medium petri plates to protect the integrity of both needles and embryos. Throughout this process the embryos were kept hydrated in PTW.

According to the developmental stage of each set of embryos, embryos were digested with proteinase K (10μg/ml) in PTW for different periods of time.

Embryos were immediately refixed in 4% PFA in PTW for 20 minutes at room temperature, after proteinase K digestion, and then washed five times with PTW, with each wash lasting five minutes.

Embryos were incubated in Hybridization Mix (Hybmix) [50% formamide (Roche); 5xSSC; 0,1% Tween 20; citric acid to pH 6,0; 50μg/ml heparin; 500μg/ml tRNA] between 3 and 4 hours at 70ºC, prior to overnight incubation with probe at 70ºC, having this one been previously diluted in Hybmix and heated.

In the next day, the probe solution is removed from the embryos and stored at -20ºC. Next, the embryos are washed at 70ºC in 100%Hybmix, 25% 2xSCC in Hybmix, 50% 2xSCC in Hybmix, 75% 2xSCC in Hybmix and 100% 2xSCC, with all washes lasting 15 minutes with the exception for the first one which lasted 10 minutes. Embryos were then washed again, this time at room temperature, in 0,2x SCC, two times with each lasting 15 minutes, 50% 0,2x SSC in PTW for 10 minutes and finally, two times in PTW with each lasting 10 minutes.

Embryos were incubated in blocking solution [2% sheep serum; 2mg/ml BSA in PTW] at room temperature for approximately two hours and then a 1:5000 dilution of Anti-Digoxigenin-AP (Roche) in blocking solution was used to incubate the embryos overnight.

In the last day, the embryos are washed 15 minutes for six times in PTW, and then 3 times in Staining Buffer [100mM TrisHCL pH 9,5; 50mM MgCl2; 100mM NaCL; 0,1% tween 20] for five minutes each.

Then, the probe was revealed by incubating the embryos in purple AP substrate in the dark at room temperature. Revelation was stopped by removing the substrate in exchange with PTW. Embryos were then fixed once more in PFA for 20 minutes at room temperature, and sequentially transferred to 100% PTW, 20% glycerol (sigma) in PTW, 50% glycerol in PTW and 80% glycerol. Embryos were stored at 4ºC.

2.12 High Resolution Melting

High Resolution Melting (HRM) was performed on Rotor-Gene 6000 (Corbett Life Science), using 2x Power SYBR Green PCR Master Mix (Applied Biosistems). Reaction conditions were optimized during the course of this work. The primers and amplicons were designed according to Rotor-Gene 6000 assay guide, using both NCBI blast and the DINAMelt Servers from the Rensselaer Polytechnic Institute. Primers were also designed with similar melting temperatures so that one single reaction condition could be used in all three situations. Since 2x Power SYBR Green Master Mix is best suited for Applied Biosystems Real-Time PCR Systems according to the manufacturer's instructions, a search was done to try to find in the literature cases where Power SYBR and Rotor-Gene 6000 were used together. Both (Pornprasert et al. 2008) and (Price et al. 2007) proved consistent. Moreover, (Pornprasert et al. 2008) used successfully a three step PCR, which was important since the primers used here did not have high melting temperatures.

DNA amplification was performed in 20μl reaction volume containing: 10μl of 2x Power SYBR Green Master Mix, 0,4μM of each primer and 1μl of DNA sample. The reaction conditions included a pre-heat at 95ºC for 10 minutes to activate AmpliTaq Gold DNA Polymerase (included in the Master Mix), and then 45 cycles of 95ºC for 20 seconds, 55ºC* for 20 seconds and 72ºC for 45 seconds. HRM curve was obtained, following DNA amplification, by heating samples for 65ºC to 95ºC at a rate of 0.1ºC per 2 seconds.

HRMA was also done using Rotor-Gene 6000 software.

* for primer pair "*HRM_dmrt2a_pair2*" 56ºC were used.

2.13 Mutant zebrafish line

After carefully dealing with all the in vitro aspects of assembling the constructs and synthetizing the mRNA molecules, a distinct part of this work is initiated, this one regarding the *in vivo* aspects of making a TALEN mutant zebrafish line. The first step is to optimize the amount of mRNA to be injected into one cell stage embryos. From as little as 4pg per embryo (Dahlem et al. 2012), to 600pg (Sander et al. 2011), the correct amount of mRNA to be injected has to be defined, one that retrieves a significant number of embryos targeted by the TALENs but one that does not compromise neither the specificity of the cut, or the amount of embryos that die due to toxicity.

Already referred is the range of different cuts that TALENs can produce. The embryos injected with TALENs mRNA at one cell stage, may or may not acquire mutations. Moreover, these mutations can be different from cell to cell, giving rise to mosaics, and may not even be present in the germline. To overcome this, a number of different zebrafish mosaics for the same kind of mutation should be crossed with wild type ones, in order to evaluate the presence of germline transmission and also to assess the different alleles generated by the TALENs. This generation, called F1 generation is heterozygous, possessing one mutant allele from a mosaic zebrafish, and a wild type allele from a wild type zebrafish. It is this F1 heterozygous generation that can be incrossed in order to produce a homozygous mutant, at a 25% rate.

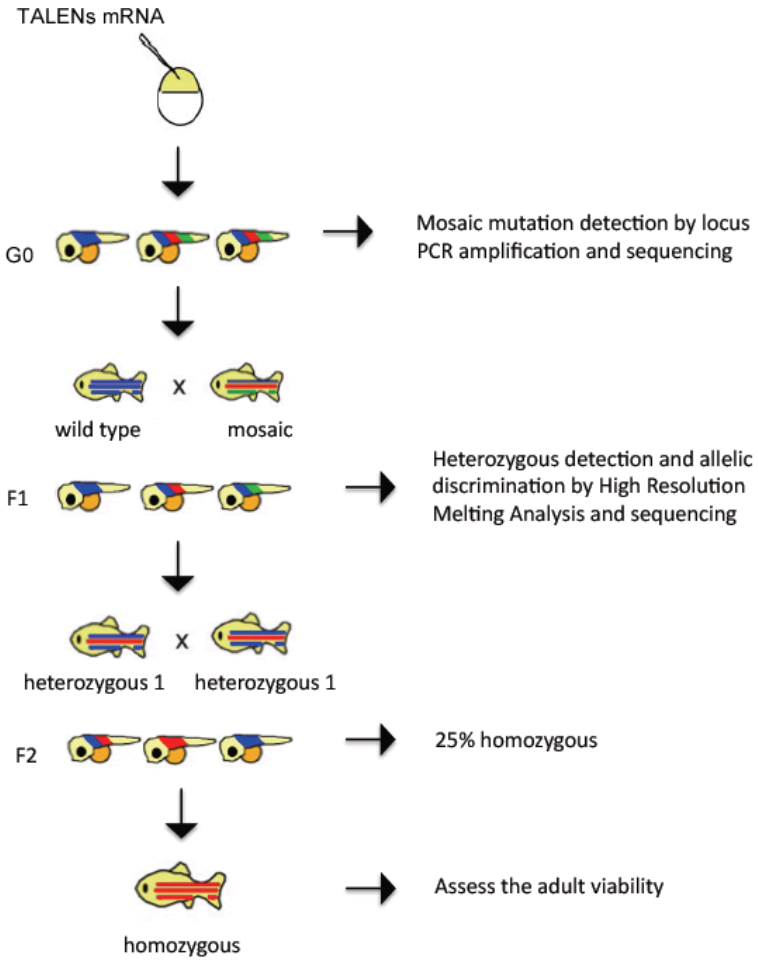


Figure 2.1 Diagram of the zebrafish cross plan to develop a mutant line. Upon injection of Transcription activator-like effectors nucleases (TALENs) mRNA at the one cell stage, G0 embryos are sacrificed and genotyped to assess which ones possess mosaic mutations. The G0 mosaic fish are then crossed with wild type ones, to generate an F1 heterozygous generation carrying only one mutated allele. F1 heterozygous fish are once again genotyped and the ones possessing the same exact mutated allele can be incrossed in order to generate an F2 homozygous mutant. Adapted from (Kawakami 2005).

Through all these processes, from the injection of TALENs mRNA to the generation of a homozygous mutant zebrafish, the most important aspect is mutation detection and discrimination. After TALENs injection, fish have to be genotyped to assess the efficiency of the TALENs cut and then germline transmission has to be assessed with the heterozygous fish being sorted considering the mutant allele that was acquired. During this work this is achieved by locus PCR amplification and sequencing and High Resolution Melting analysis (HRMA).

Chapter 3

Results

3.1 Generating zebrafish *dmrt2a* and *dmrt2b* mutant alleles with TALENs

3.1.1 Deciding on the TALEN target sequence

The first step towards the design of TALEN pairs consisted on deciding where to target the genes *dmrt2a* and *dmrt2b.* In Figure 3.1 a general architecture of these two genes is represented. The objective was to design two TALEN pairs for each gene, one that would target the ATG region and another that would target a region just before the start of the DM domain, thought to be the domain with which Dmrt2a and Dmrt2b proteins bind the DNA (Murphy et al. 2007). The first TALEN pair would be used to try to develop null mutants, and the second TALEN pair to try to create shift mutations that disrupted the reading frame.

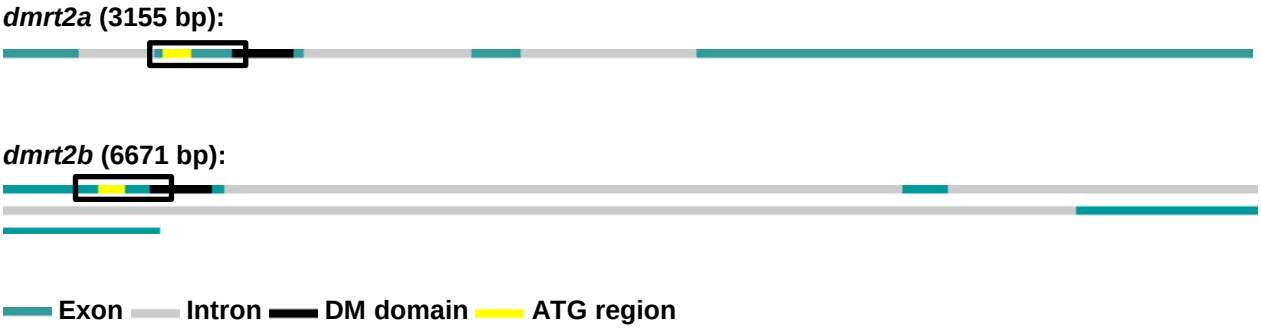


Figure 3.1 Representation of *dmrt2a* and *dmrt2b* genomic sequences. *dmrt2a* is composed of 3155 bp, with 4 exons (blue) and 3 introns (grey). The ATG and the DM domain of *dmrt2a* are situated in the second exon distancing 171 bp from each other (upper black rectangle). *dmrt2b* is composed of 6671 bp, with 3 exons (blue) and 3 introns (grey). The ATG and the DM domain of *dmrt2b* are situated in the first exon and distance 138 bp from each other (lower black rectangle).

To ensure that the presence of possible polymorphisms in the selected regions (ATG and before the DM domain) of our zebrafish wild type lines would not affect the binding efficiency of TALENs, 10 wild type TU adult zebrafish were genotyped. The zebrafish adult fish with no relevant polymorphisms were used to generate the embryos where TALENs were injected. Represented as black rectangles in Figure 3.1 are the regions that were sequenced during this process. They compromise the ATG region and the start of the DM domain.

A

```
wt TU 1    118 AGTAAAATGACAGCATTTAACTTGACACGTTACATGCAGGAAAACATAGTTTTAAAAAAA 177
wt TU 2    118 ............................................................ 177
wt TU 3    117 ............................................................ 176
wt TU 4     84 ...........................................A................ 143
wt TU 5    118 ............................................................ 177
wt TU 6    116 ...........................................A................ 175
wt TU 7    117 ...........................................A................ 176
wt TU 8    117 ...........................................A................ 176
wt TU 9    117 ...........................................A................ 176
wt TU 10   117 ...........................................A................ 176
                                                        ATG
wt TU 1    178 TAATTATTCGCATTTTTTTGGACAGCATCTTGCTTTACGCGGCCAGAAAATGACGGATCT 237
wt TU 2    178 ............................................................ 237
wt TU 3    177 ............................................................ 236
wt TU 4    144 ............................................................ 203
wt TU 5    178 ............................................................ 237
wt TU 6    176 ............................................................ 235
wt TU 7    177 ............................................................ 236
wt TU 8    177 ............................................................ 236
wt TU 9    177 ............................................................ 236
wt TU 10   177 ............................................................ 236

wt TU 1    238 GTCCGGCACGGAGTTTGAGATCGACGTGGAGGGCCTGGAGACGGAGAGCGATGATCAGGC 297
wt TU 2    238 ............................................................ 297
wt TU 3    237 ............................................................ 296
wt TU 4    204 ............................................................ 263
wt TU 5    238 ............................................................ 297
wt TU 6    236 ............................................................ 295
wt TU 7    237 ............................................................ 296
wt TU 8    237 ............................................................ 296
wt TU 9    237 ............................................................ 296
wt TU 10   237 ............................................................ 296

wt TU 1    298 GGTGTTCACCGGCCCCGGAGAGGACGACACGGGGTCCAAAGACGACGACAAATCTGACCG 357
wt TU 2    298 ............................................................ 357
wt TU 3    297 ............................................................ 356
wt TU 4    264 ............................................................ 323
wt TU 5    298 ............................................................ 357
wt TU 6    296 ............................................................ 355
wt TU 7    297 ............................................................ 356
wt TU 8    297 ............................................................ 356
wt TU 9    297 ............................................................ 356
wt TU 10   297 ............................................................ 356
                                                          DM domain
wt TU 1    358 CTTAAATCTCGGCTCTGGAGACCAGCGCAAACTGAGCCGCACGCCGAAATGCGCCCGATG 417
wt TU 2    358 ............................................................ 417
wt TU 3    357 ............................................................ 416
wt TU 4    324 ............................................................ 383
wt TU 5    358 ............................................................ 417
wt TU 6    356 ............................................................ 415
wt TU 7    357 ............................................................ 416
wt TU 8    357 ............................................................ 416
wt TU 9    357 ............................................................ 416
wt TU 10   357 ............................................................ 416
```

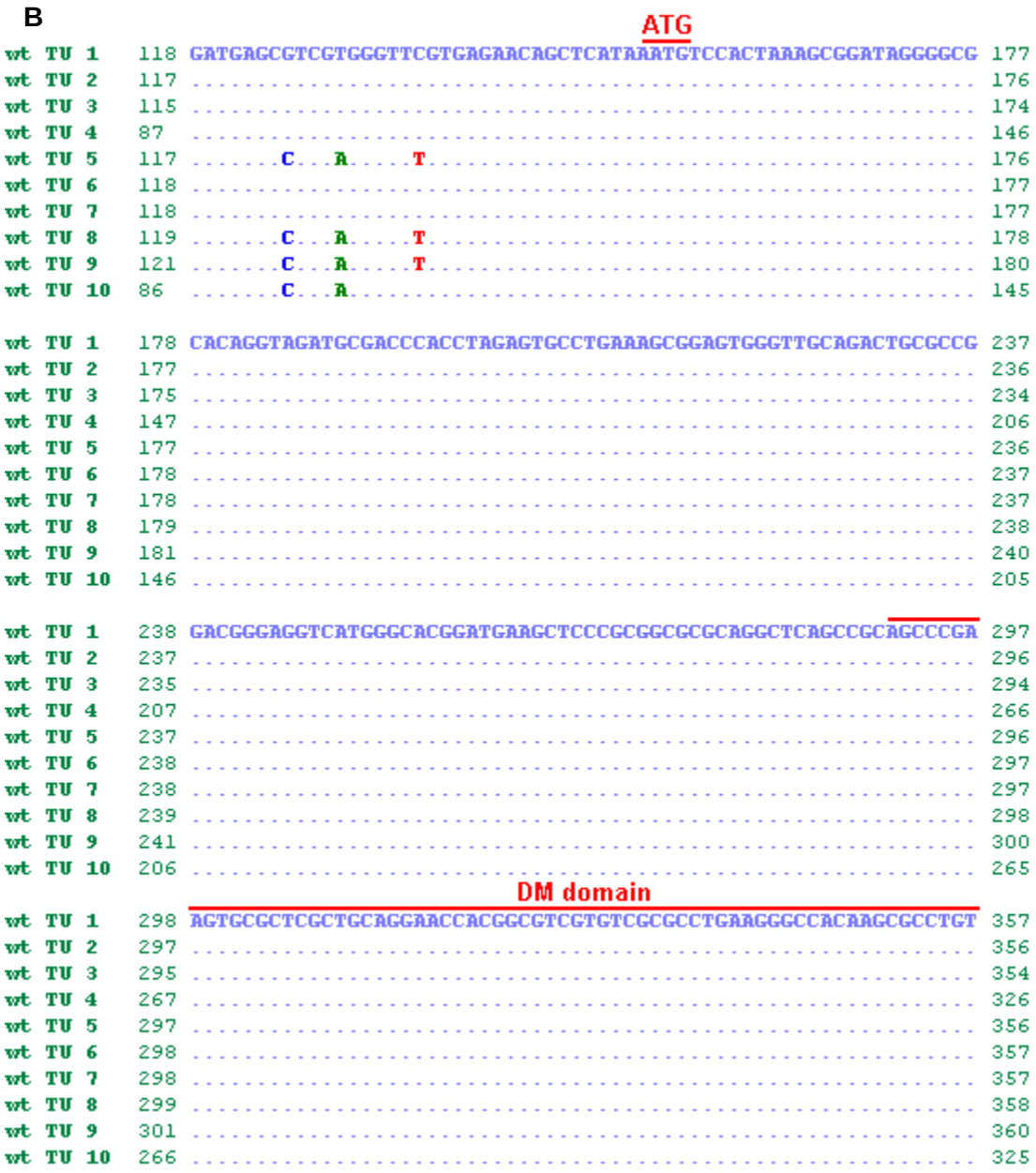


Figure 3.2 Wild type adult zebrafish genotyping. (A, B) 10 wild type adult zebrafish were genotyped for *dmrt2a* (panel A) and *dmrt2b* (panel B). For both genes, the DM domain extends further than the sequenced region. No polymorphisms were found for both genes in between the ATG and the DM domain. (A) The genotyped region of *dmrt2a* only revealed one polymorphism (A/T) upstream of the ATG. (B) The genotyped region of *dmrt2b,* revealed several polymorphisms upstream of the ATG.

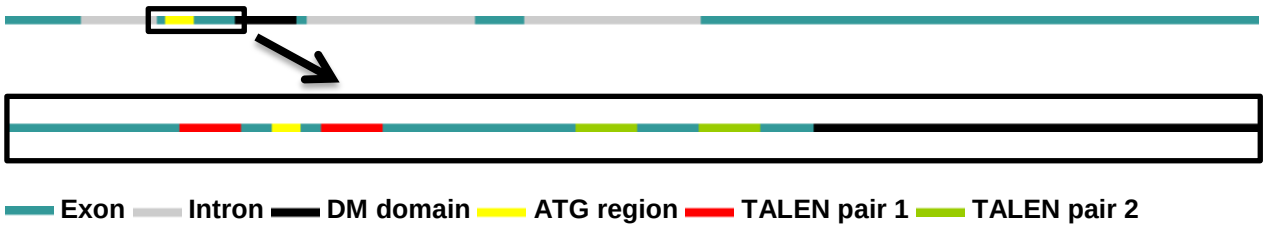
For the gene *dmrt2a*, only one polymorphism (A/T) was detected upstream of the ATG region but at a sufficient distance that would not affect the desired binding site of the first *dmrt2a* TALEN pair (panel A of Figure 3.2). For the gene *dmrt2b* several polymorphisms were found upstream of the ATG region exactly at the site where one would normally design a TALEN monomer to target the ATG region (panel B

of Figure 3.2). For both genes, no polymorphisms in between the ATG region and the DM domain were found.

3.1.2 TALENs design and assembly

Four TALEN pairs were designed and assembled as described in Materials and Methods. For *dmrt2a* one TALEN pair was designed and assembled to target the ATG region (*dmrt2a*-pair1) with another TALEN pair being designed and assembled to target a region prior to the start of the DM domain (*dmrt2a*-pair2). As for the gene *dmrt2b* two partially similar TALEN pairs were designed and assembled to target a region in between the ATG and the start of the DM domain (*dmrt2b*-pair1 and *dmrt2b*-pair2). It is important to note that adding to the fact that several polymorphisms were detected prior to the ATG region of *dmrt2b* (as described in the last section), no candidate TALEN pairs would respect the NG + HD > NI + NN rule for the ATG region of *dmrt2b* (already described). A schematic view of each TALEN pair target region within the *dmrt2a* and *dmrt2b* genomic sequences can be seen below in Figure 3.3.

dmrt2a* TALENs*:



dmrt2b TALENS:

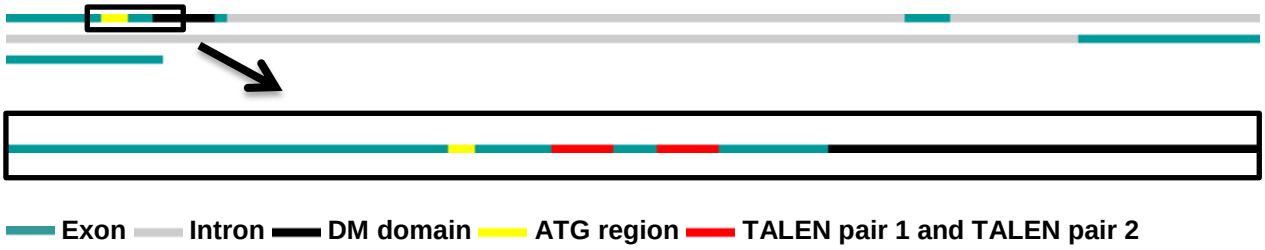


Figure 3.3 Schematic representation of each TALEN pair target region within the *dmrt2a* and *dmrt2b* genomic sequences. For both genes a magnification of the genomic region that compromises the ATG and the start of the DM domain is shown. For *dmrt2a* two TALEN pairs were designed: one pair (upper red stripes) to target the ATG region (yellow stripes) and one pair (upper green stripes) to target a region prior to the start of the DM domain (black stripes). For *dmrt2b* also two TALEN pairs were designed but these target the same region in between the ATG and the DM domain, and so are represented together in the lower red stripes.

The assembling process of TALENs consists of repetitive cloning steps where each RVD sequence is joined together in a specific order that confers binding specificity to each TALEN monomer. In Figure 3.4 these RVD sequences are represented for each TALEN monomer.

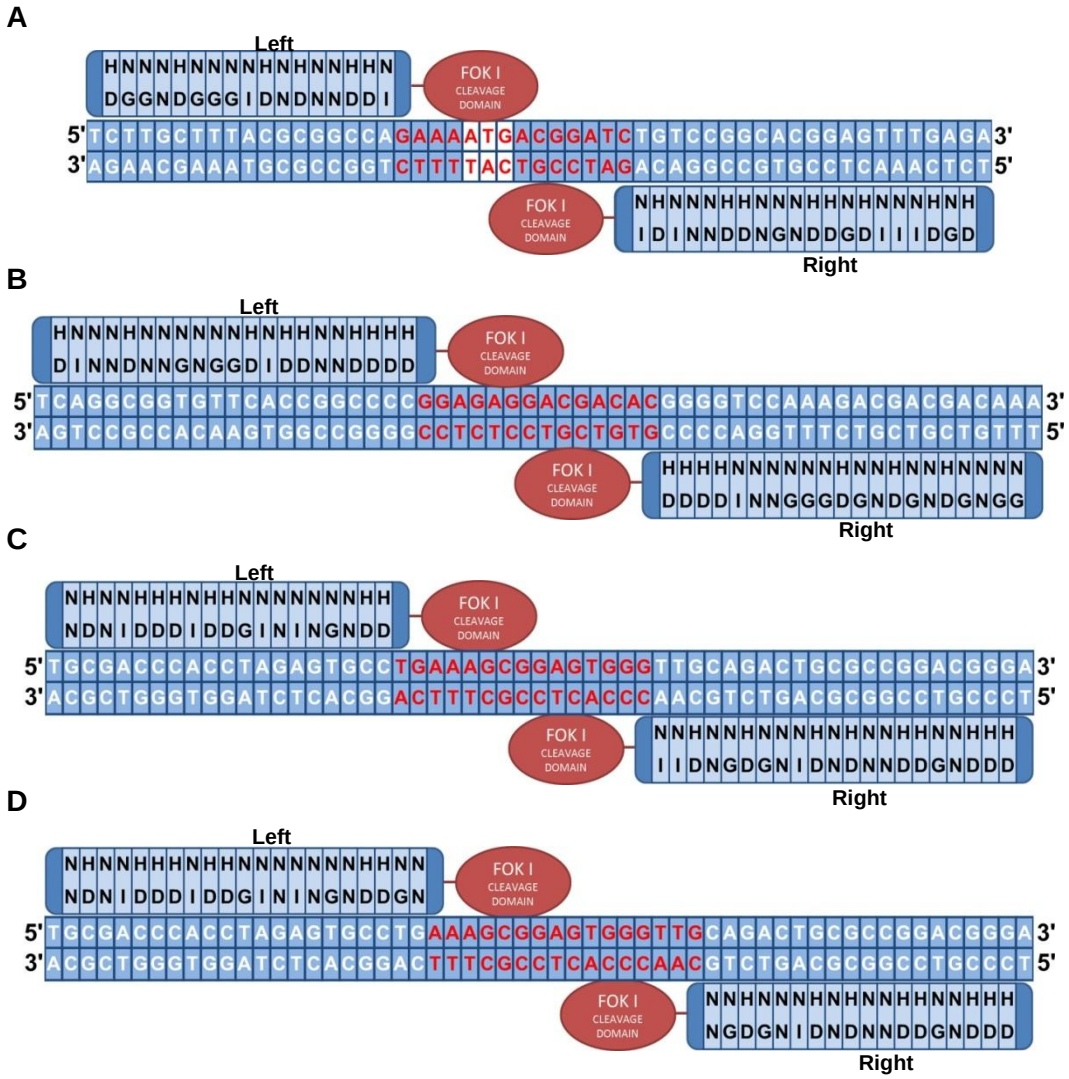


Figure 3.4 RVD sequence representation for each TALEN monomer assembled during this work. (A) Left and Right monomers of *dmrt2a*-pair1 with the ATG highlighted in white. (B) Left and right monomers of *dmrt2a*-pair2. (C) Left and right monomers of *dmrt2b*-pair1. (D) Left and right monomers of *dmrt2b*-pair2.

As can be seen in panel A of Figure 3.4, the target site of *dmrt2a*-pair1 includes the ATG of dmrt2a. Also perceptible from panels C and D from Figure 3.4, is the similarity between *dmrt2b*-pair1 and *dmrt2b*-pair2 monomers, with its RVD sequences only differing close to the spacer region.

3.1.3 TALENs mRNA injection (Mosaic G0 generation)

3.1.3.1 Single TALEN pair mRNA injection

In order to generate zebrafish mosaics for the *dmrt2a* and *dmrt2b* mutant alleles, previously genotyped zebrafish adults were incrossed, and embryos at the one cell stage were microinjected with TALENs mRNA. As described in Materials and Methods, mRNAs encoding left or right monomers were individually *in vitro* synthesized from linearized plasmids, but co-injected together in equal amounts. For each TALEN pair injected, groups of embryos with approximately one week post fertilization were sacrificed so that their genomic DNA could be extracted and sequenced to assess the presence of mosaic mutations.

In a first attempt to induce genomic DNA sequence alterations in the gene *dmrt2a*, embryos were injected with 50pg of *dmrt2a*-pair1 or *dmrt2a*-pair2 mRNAs. Sequencing results representative of these first attempt can be observed in panels A of both Figures 3.5 and 3.6. As for panels B, these show sequencing results representative of a second attempt where 200pg of *dmrt2a*-pair1 or *dmrt2a*-pair2 mRNAs were injected.

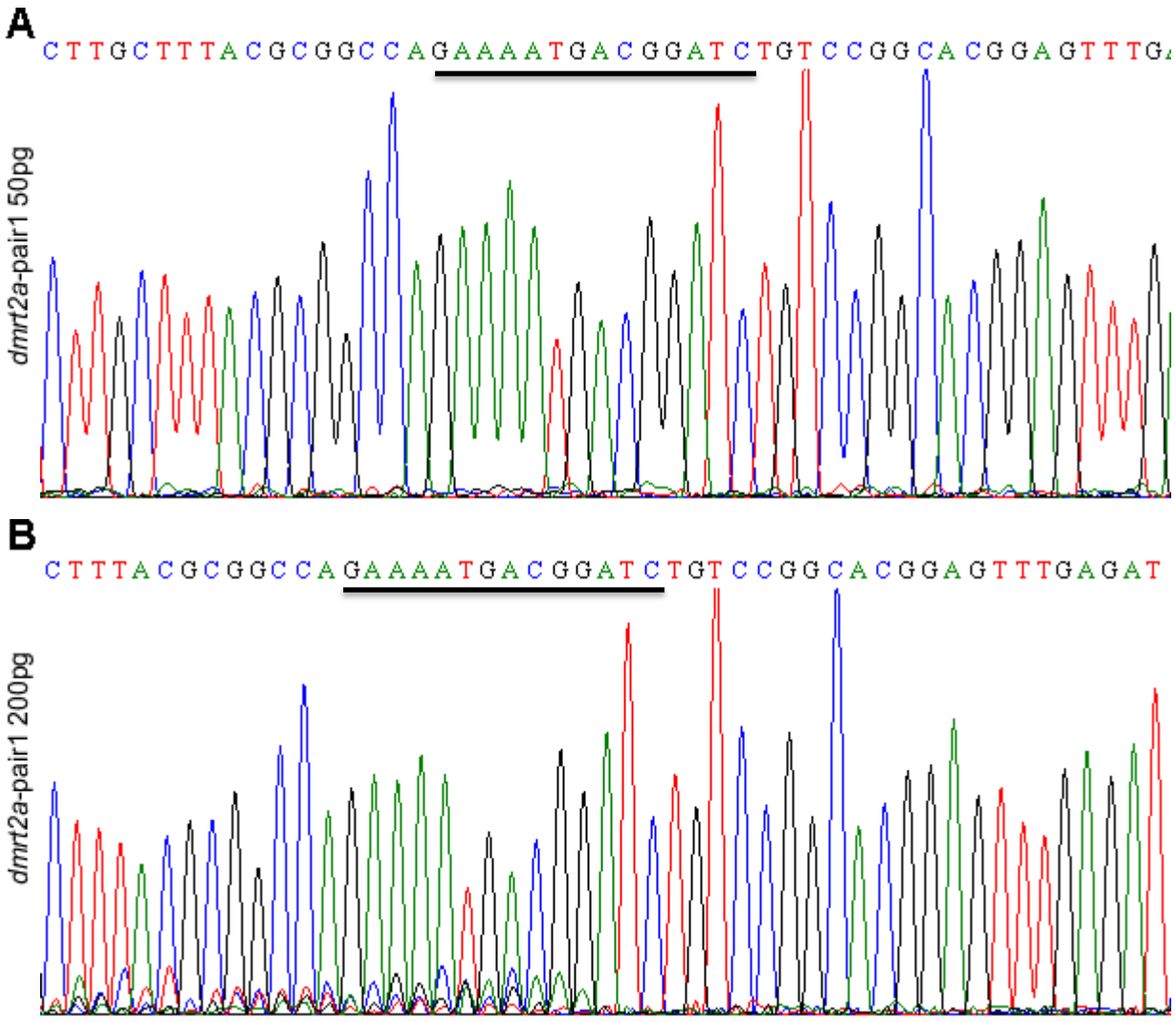


Figure 3.5 Comparison between the injection of 50pg and 200pg of *dmrt2a*-pair1 mRNA. (A) genomic DNA from a 50pg *dmrt2a*-pair1 mRNA injected embryo sequencing result. (B) genomic DNA from a 200pg *dmrt2a*-pair1 mRNA injected embryo sequencing result. Only when 200pg of *dmrt2a*-pair1 mRNA are injected, multiple peaks appear at the target site (panel B). Target site is underlined.

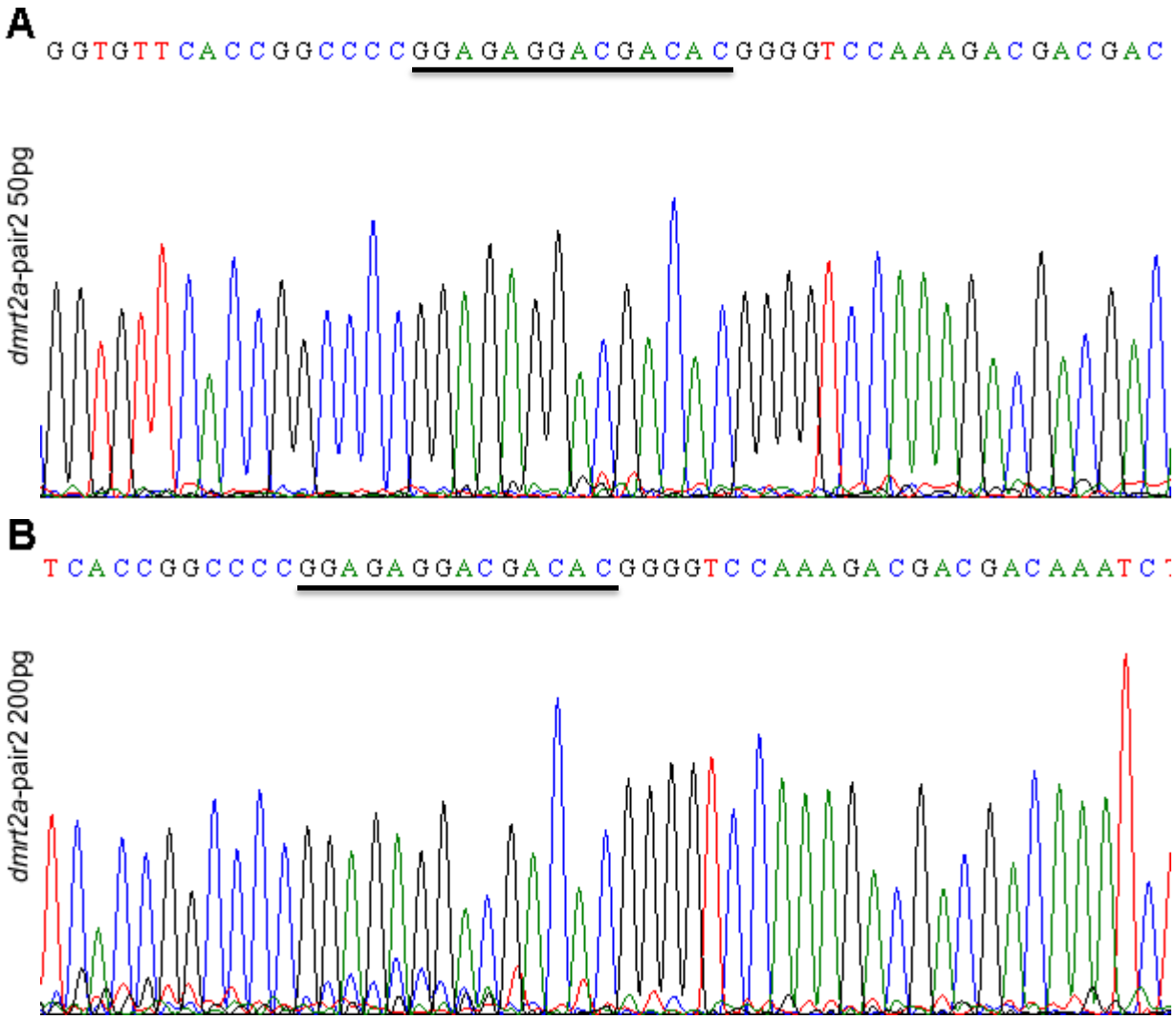


Figure 3.6 Comparison between the injection of 50pg and 200pg of *dmrt2a*-pair2 mRNA. (A) Genomic DNA from a 50pg *dmrt2a*-pair2 mRNA injected embryo sequencing result. (B) genomic DNA from a 200pg *dmrt2a*-pair2 mRNA injected embryo sequencing result. Only when 200pg of *dmrt2a*-pair2 mRNA are injected, multiple peaks appear at the target site (panel B). Target site is underlined.

When 50pg of *dmrt2a*-pair1 or *dmrt2a*-pair2 mRNAs were injected, no mosaic peaks could be observed (panels A of Figures 3.5 and 3.6). This does not mean that the amount of TALENs injected was not sufficient to induce DNA sequence alterations. It could just be that the number of mosaic mutations present within a given embryo was not enough to be detected by the sequencing reaction. Nevertheless, when 200pg of *dmrt2a*-pair1 or *dmrt2a*-pair2 mRNAs were injected, multiple peaks could be observed at the TALENs target site (underlined in each panel) for both the first and second pairs of *dmrt2a* TALENs (panels B of Figures 3.5 and 3.6). It is important to note that all the sequencing reactions present in this section were obtained with the respective reverse primers which mean that the mosaic peaks should span the entire left side of each panel.

With the first results already indicating that the amount of injected TALENs should be around 200pg per embryo, it was decided to inject 250pg per embryo of the *dmrt2b*-pair1 (Figure 3.7). *dmrt2b*-pair2 (Figure 3.8) had to be injected with 200pg and 400pg per embryo (explained further ahead).

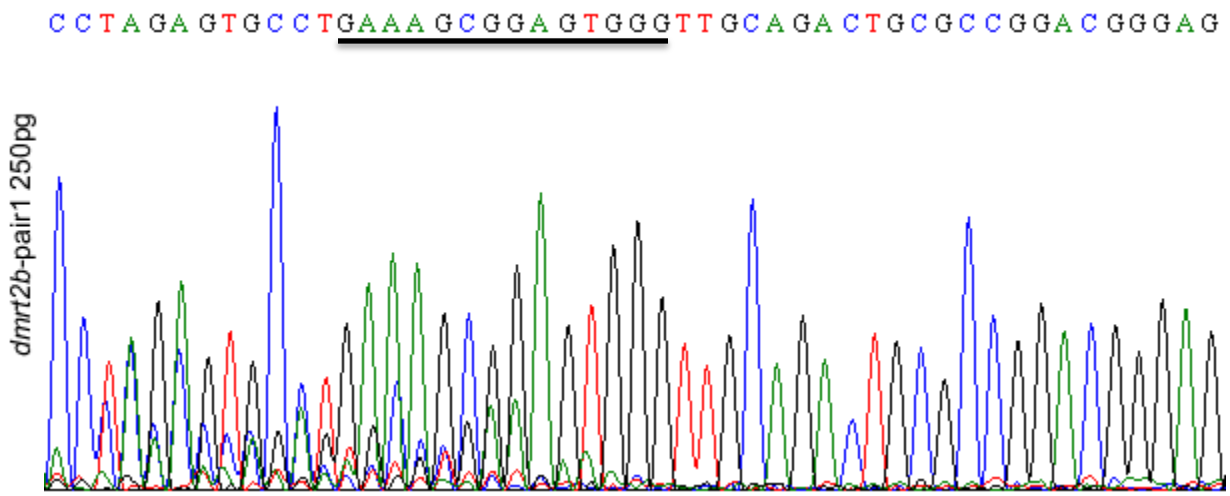


Figure 3.7 Injection 250pg of *dmrt2b*-pair1 mRNA. Genomic DNA from a 250pg *dmrt2b*-pair1 mRNA injected embryo sequencing result. Multiple peaks appear at the target site (underlined).

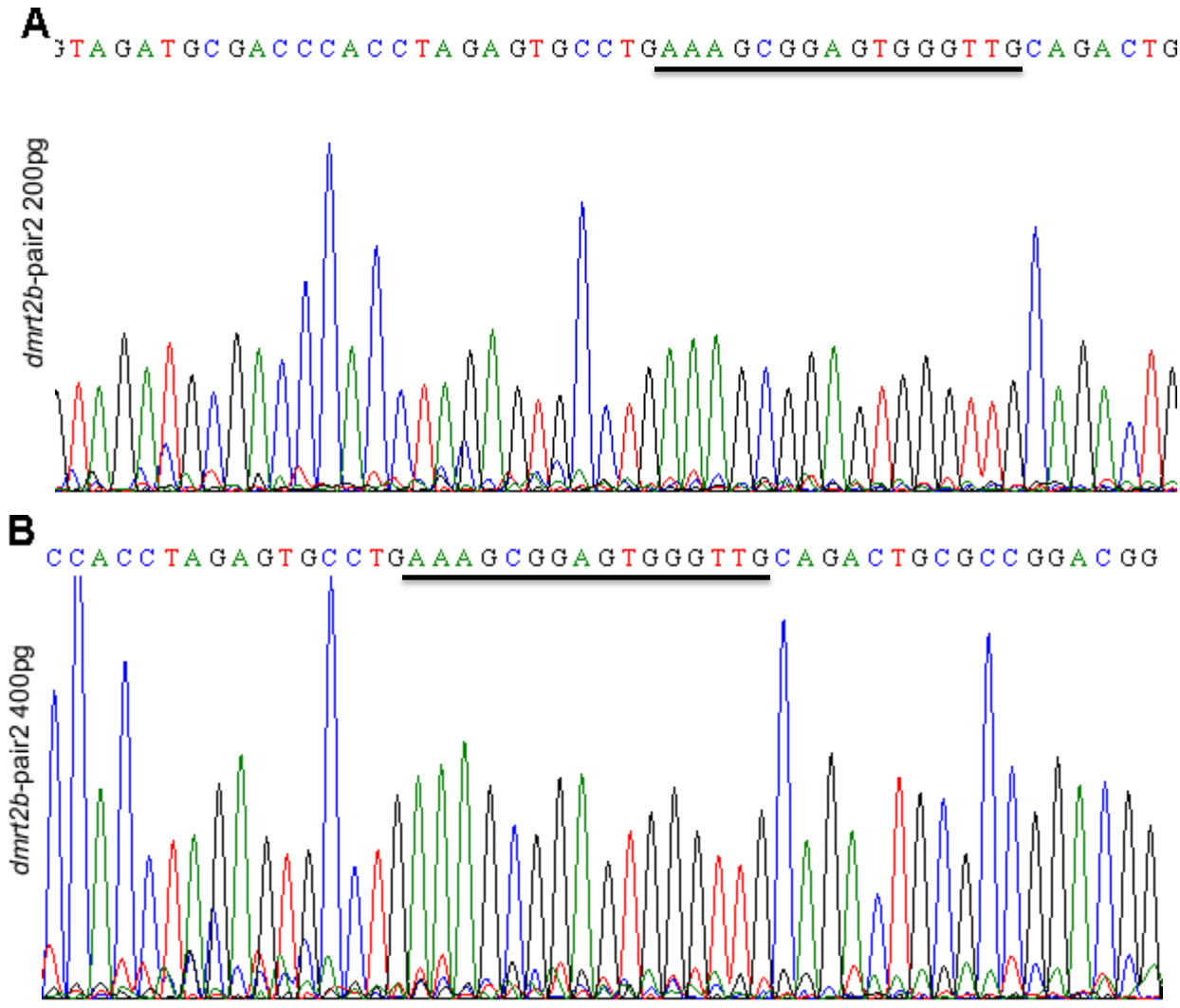


Figure 3.8 Comparison between the injection of 200pg and 400 pg of *dmrt2b*-pair2 mRNA. (A) Genomic DNA from a 200pg *dmrt2b*-pair2 mRNA injected embryo sequencing result. (B) Genomic DNA from a 400pg *dmrt2b*-pair2 mRNA injected embryo sequencing result. Only when 400pg of *dmrt2b*-pair2 mRNA are injected, multiple peaks appear at the target site (panel B). Target site is underlined.

As described in the previous section, these two *dmrt2b* TALEN pairs were structurally similar, targeting the exact same region in the gene, with only 3 bp difference. It was expected that a similar amount of TALENs mRNA would provide the same results but this was not the case. 250pg of *dmrt2b*-pair1 were enough to induce genomic DNA sequence alterations in a number that some of the mosaic peaks were as high as the wild type ones (Figure 3.7). As for the second *dmrt2b* TALEN pair, only the injection of 400pg of mRNA was sufficient to induce clear mosaic mutations as seen in panel A and B of Figure 3.8. Also, when injecting 400pg of *dmrt2b*-pair2 mRNA it seems that the mosaic peaks span a region that goes beyond the target site, as can be observed in panel B of Figure 3.8. Unlike the previous

situations, here it is possible to observe multiple peaks at the right side of the TALEN target site possibly indicating that the DNA alterations induced were bigger.

Table 3.1 Summary of single TALEN pairs injection. Table 3.1 shows the different amounts of TALEN pairs mRNA that were injected as well as the number of mosaic embryos detected in each case.

TALENs injected	sequenced embryos	mosaic embryos	% of mosaic embryos
***dmrt2a* pair 1 50pg**	10	1	1
***dmrt2a* pair 1 200pg**	10	9	90
***dmrt2a* pair 2 50pg**	10	0	0
***dmrt2a* pair 2 200pg**	10	9	90
***dmrt2b* pair 1 250pg**	5	5	100
***dmrt2b* pair 2 200pg**	6	3	50
***dmrt2b* pair 2 400pg**	7	6	86

3.2.3.2 Multiple TALEN pair mRNA injection

Recently it has been described that TALENs have the ability to create big lesions (more than 100 bp) when multiple TALEN pairs are used together (Xiao et al. 2013). As already described during this work four TALEN pairs were designed. Although both *dmrt2b* TALEN pairs target the same site in the genome, TALENs for *dmrt2a* have its target sites separated by 100 bp. It was then decided to test the activity of this two TALEN pairs simultaneously. Both TALEN pairs were co-injected in a total concentration of 400pg per embryo. A group of embryos was then sacrificed so that their genomic DNA could be individually extracted and sequenced to test TALENs activity.

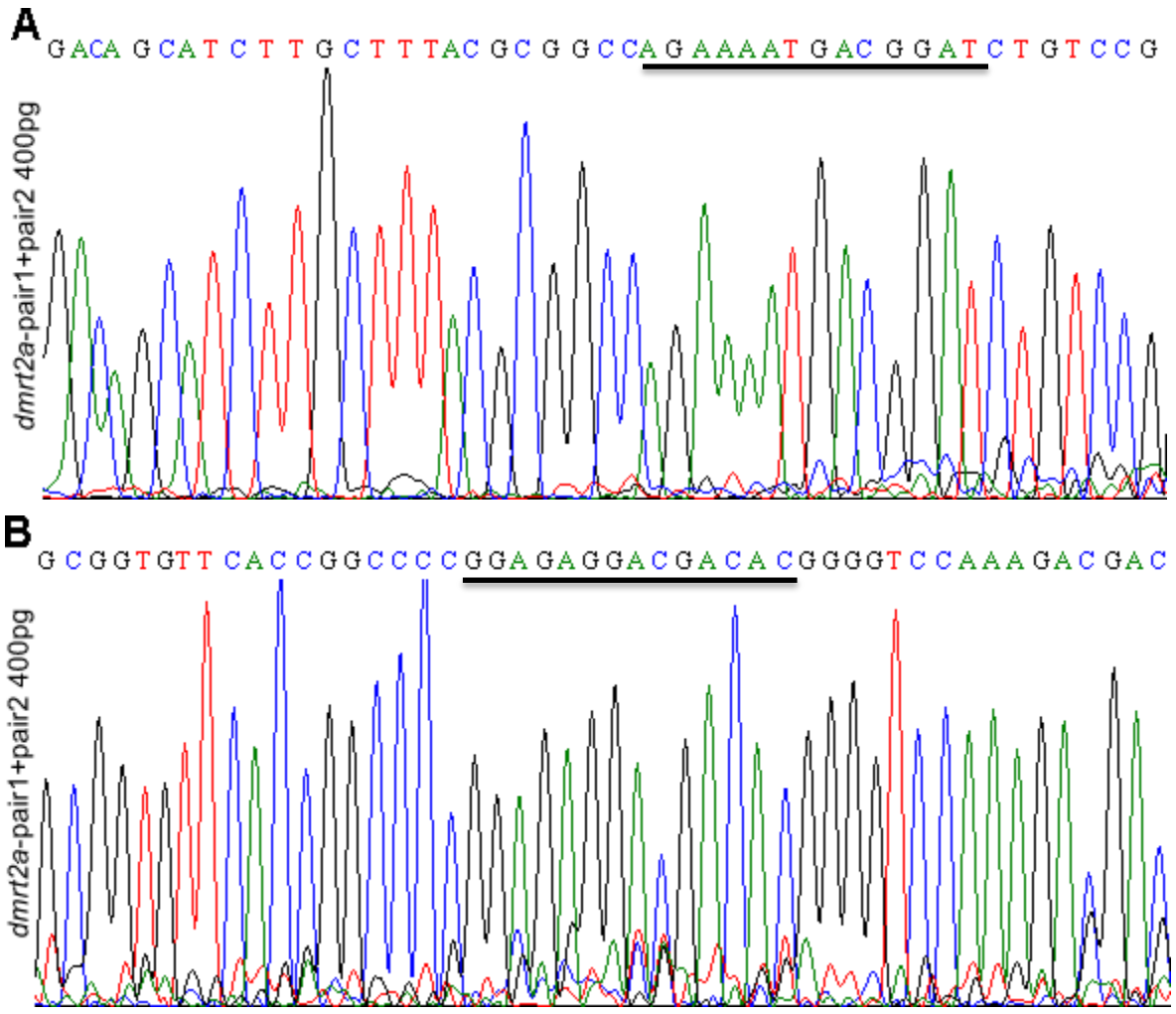


Figure 3.9 Sequencing results from the injection of *dmrt2a*-pair1 and *dmrt2a*-pair2 simultaneously. (A) The sequencing reaction is performed with a forward primer to assess the presence of mosaic peaks at the ATG region. (B) The sequencing reaction is performed with a reverse primer to assess the presence of mosaic peaks at the target site of *dmrt2a*-pair2. Each pair has its target site underlined.

As expected, at the target site of *dmrt2a*-pair1 multiple peaks could be observed (panel A Figure 3.9), but that was not clear in respect to the target site of *dmrt2a*-pair2 (panel B Figure 3.9). It was not easy to understand if simply no lesions were made with *dmrt2a*-pair2 (possibly the multiple peaks observed were due to errors in the sequencing reaction) or if the actual peaks observed in panel B of Figure 3.9 resembled the mosaic peaks from panel B of Figure 3.8 where the lesion made with TALENs spanned a region that went much beyond the target site.

Until here, the ability of TALENs to generate mosaic mutations had been confirmed only by sequencing the target site with a reverse primer expecting to detect multiple peaks starting at the TALEN target site. In this case, and if the objective was to create at least a 100 bp lesion, that proved not to be sufficient. Sequencing the region with primers flanking the two target sites would only give information on whether TALENs were working or not, without giving information on possible lesions in between the two

cut sites. PCR primers designed specifically to be used in the next section of this thesis, were adapted and used also in this step. Together, primers HRM_*dmrt2a*_pair1_FW and HRM_*dmrt2a*_pair2_RV amplify a 200 bp fragment spanning the two *dmrt2a* TALEN pairs cut sites. The same samples that had already been sequenced were then used in a PCR reaction to amplify this 200 bp fragment. Also, it was used as negative control, a genomic DNA sample from a zebrafish embryo that had not been exposed to TALENs.

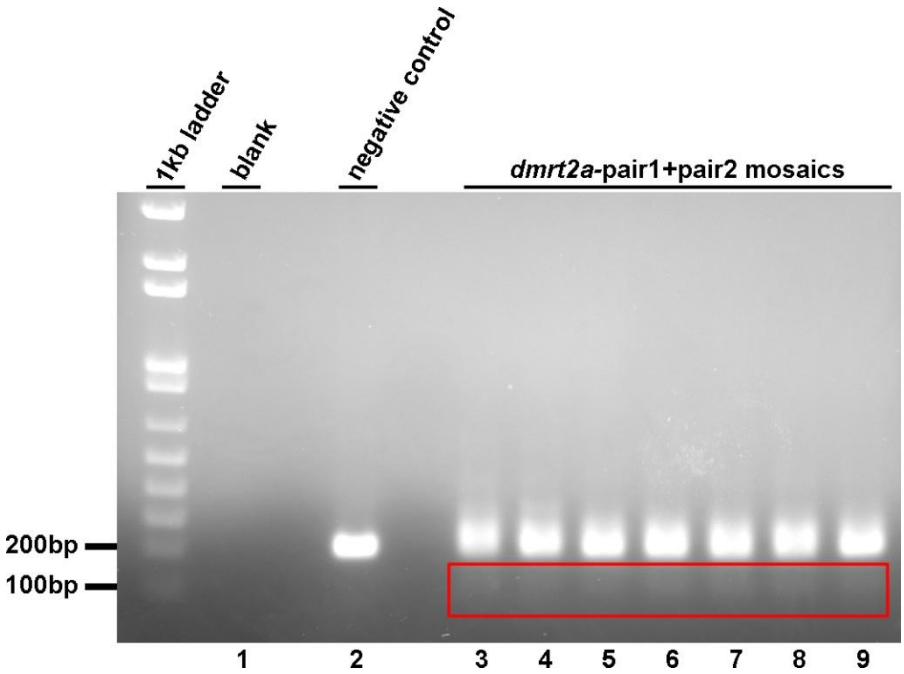


Figure 3.10 Electrophoresis run of *dmrt2a*-pair1+pair2 samples reveal the presence of faint bands below 200 bp. Genomic DNA samples from mosaic fish that were injected with *dmrt2a*-pair1+pair2 were used to amplify a fragment of 200 bp spanning the two target sites (distancing 100 bp from each other). As a negative control, a genomic DNA sample from an embryo that had not been exposed to TALENs was used (lane 2). Faint bands of approximately 100 bp are visible only in lanes 3 to 9. These correspond to the genomic DNA samples from fish injected with the two TALEN pairs simultaneously, revealing the possibility that some fragments are lacking 100 bp.

The 200 bp fragment was successfully amplified both in the negative control and in the samples from embryos exposed to the two TALEN pairs. Although difficult to see in Figure 3.10, inside the red rectangle faint bands around 100 bp can be observed only in the samples from the embryos exposed to the two TALEN pairs. Since these faint bands could not be primer dimers as they could not be observed in the blank sample, the total content of each PCR reaction was again run on an agarose gel, with the portions of gel containing this faint bands being extracted (Figure 3.11), purified and used to reamplify these faint bands with the same primer set. Samples were then run on a 4% low melting agarose gel to clearly separate any small fragments that could be amplified in the PCR reaction (Figure 3.12).

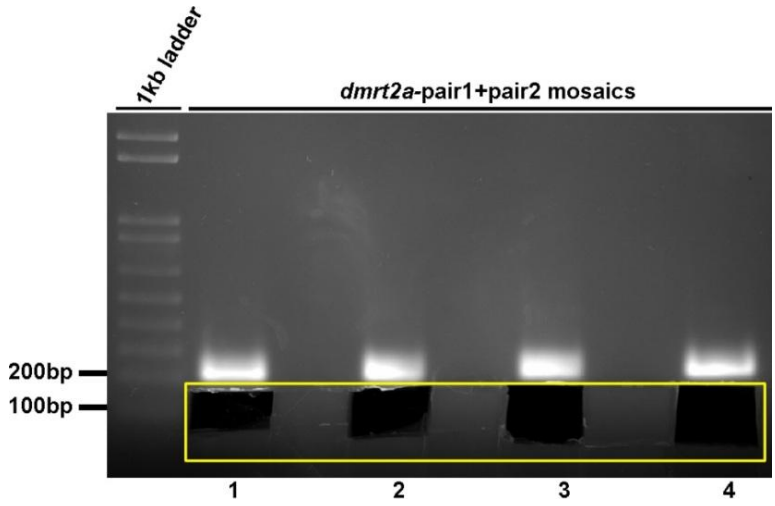


Figure 3.11 Extraction of portions of agarose gel below 200 bp. The full amplification content of 4 of the previously amplified products was run on an agarose gel. Portions of agarose gel were extracted from below the 200 bp fragments to assess if they consisted on fragments where a big genomic deletion had occurred. These were purified and then reamplified.

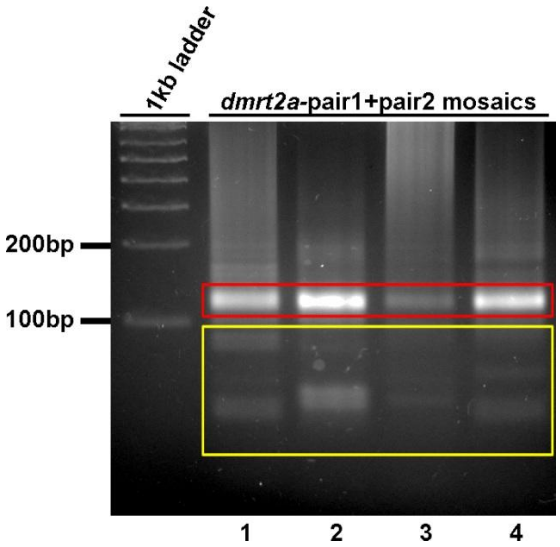


Figure 3.12 PCR reactions run on a 4% low melting agarose gel reveal the possible existence of big genomic DNA lesions. The 4 samples that were previously reamplified were run on a 4% low melting agarose gel so that any existing small fragments could be separated. 100 bp fragments were clearly visible on all 4 amplification reactions (red rectangle). These could possibly belong to fragments lacking the 100 bp that separate each *dmrt2a* TALEN target site. Apart from these 100 bp fragments, smaller fragments are clearly visible, as in lane 2 (yellow rectangle). These may belong to genomic DNA fragments where TALENs introduced a genomic deletion bigger than 100 bp.

As easily observed in Figure 3.12, the reamplification step generated different sorts of fragments, as it would be expected when a mosaic genomic DNA is being amplified. Moreover the strongest bands appear at the level of 100 bp (red rectangle), exactly the size that one would predict if the lesion induced by the TALENs would span the entire space in between the two target sites. As for the small fragments (yellow rectangle), they could explain the sequencing result from panel B of Figure 3.9. If a the lesion

would actually span a region beyond the target site, than not only the sequencing result would retrieve multiple unspecific peaks beyond the target site (panel B Figure 3.9) as the amplification of this region would result in a much smaller fragment.

To test this hypothesis not only the 100 bp fragments were sequenced as some of the smaller fragments, to confirm if they actually belonged to this *dmrt2a* genomic region (Figure 3.13)

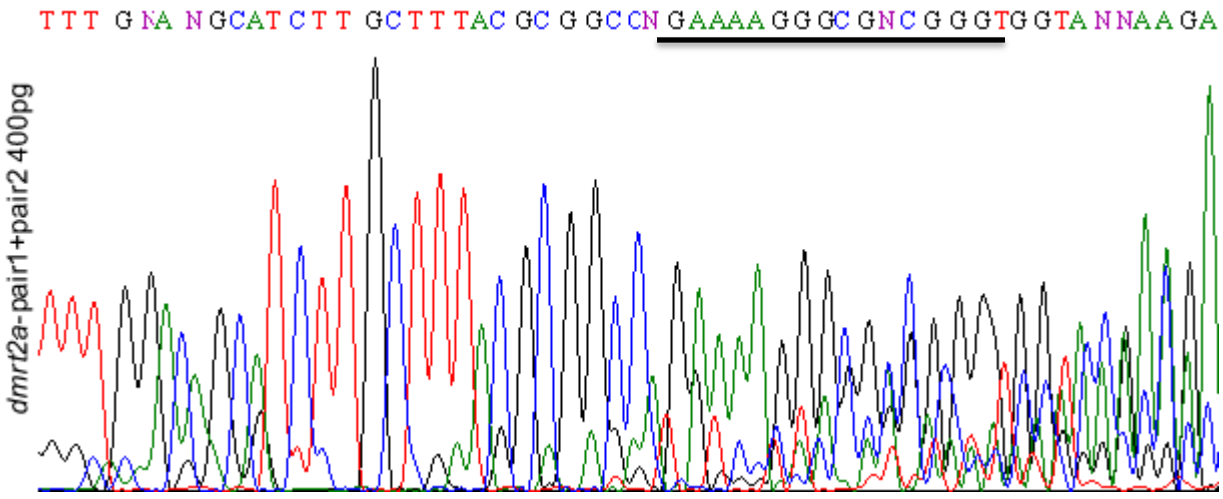
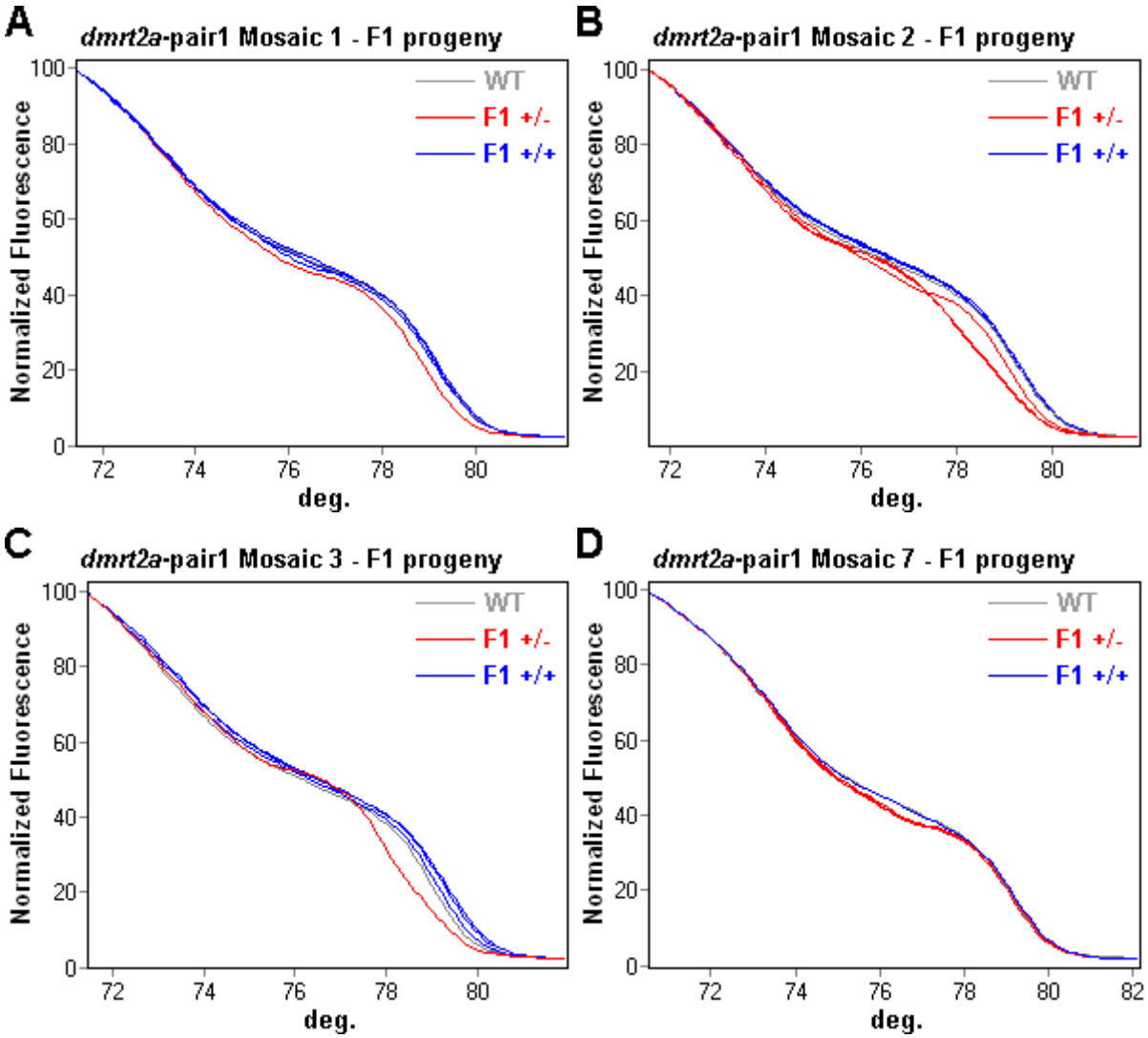


Figure 3.13 Sequencing result from a small DNA fragment amplified from genomic DNA samples of *dmrt2a*-pair1+pair2 exposed fish. A small fragment from the yellow rectangle in Figure 3.12, was extracted from the gel, purified and sequenced. The sequencing result reveals that this fragment belongs to *dmrt2a.* The sequencing reaction was performed with HRM_*dmrt2a*_pair1_FW. The nucleotides on the left of the *dmrt2a*-pair1 target site (underlined), match with *dmrt2a.*

Shown in Figure 3.13 is one of the sequencing results from the small fragments that were previously amplified. The target site of *dmrt2a*-pair1 is underlined and it was confirmed that the nucleotides before it, belong to the *dmrt2a* (panel A of Figure 3.7 can be used for comparison). As expected, mosaic peaks start appearing at the TALENs target site. This confirmed unequivocally, the ability of TALENs to be used as multiple pairs at the same time creating big DNA sequence lesions.

3.2.4 Germline transmission of acquired mutations (Heterozygous F1 generation)

TALENs injected G0 embryos from 3.1.3.1 (single TALEN pair injection) were raised to adulthood and individually crossed with wild type TU zebrafish. Embryos from each F1 progeny were individually sacrificed so that its genomic DNA could be analyzed by HRMA (Figures 3.14, 3.16, 3.18, 3.20). Genomic DNA samples that revealed HRM patterns different from the wild type ones were than sequenced so that the extent of each mutation could be assess (Figures 3.15, 3.17, 3.19, 3.21)


A
dmrt2a-pair1 Mosaic 1 - F1 progeny
B
dmrt2a-pair1 Mosaic 2 - F1 progeny
C
dmrt2a-pair1 Mosaic 3 - F1 progeny
D
dmrt2a-pair1 Mosaic 7 - F1 progeny
WT
F1 +/-
F1 +/+
Normalized Fluorescence
deg.

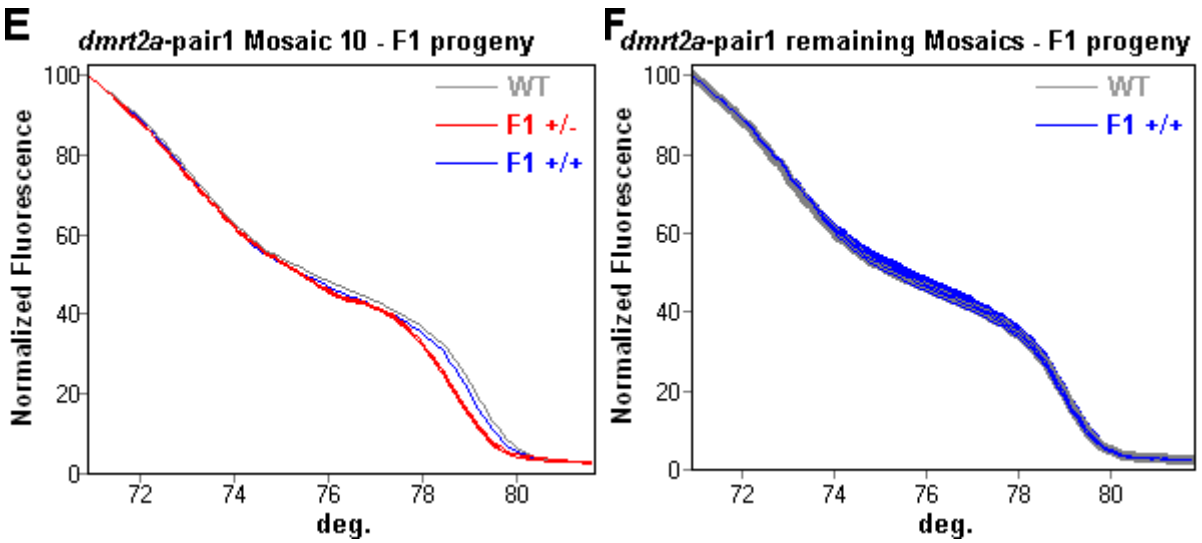


Figure 3.14 High resolution melting analysis (HRMA) of *dmrt2a*-pair1 G0 generation germline transmission. *dmrt2a*-pair1 mosaic fish were crossed with wild type ones. Individual embryos from F1 progeny were sacrificed so that genomic DNA could be extracted and HRMA could be performed. Samples from embryos not possessing any mutated allele show a melting pattern similar to the wild type (blue). The progeny of *dmrt2a*-pair1 mosaic fish that transmit mutated alleles through their germline can be assessed in panels A, B, C, D and E. In these panels, it is clear the different between the wild type melting pattern (grey) and the heterozygous ones (red). Some mosaic fish may not transmit mutant alleles through their germline and in this case the melting patterns are similar to the wild type one (panel F).

```
TCTTGCTTTACGCGGCCAGAAAATGACGGATCTGTCCGGCACGGAGTTTGAGA WT
....................---------.......................... Δ8
.................----------.......................... Δ14
....................CT-..........................  Δ3, +2
....................------.......................... Δ6
....................---------.......................... Δ9
....................-------.......................... Δ7
```

Figure 3.15 Sequenced germline transmitted mutations by *dmrt2a*-pair1 G0 generation. Samples that revealed different melting patterns from the wild type (Figure 3.14) were sequenced to confirm the presence of mutant alleles and the extent of each mutation. The binding sites of *dmrt2a*-pair1 are presented in grey. Dot: same nucleotide as the wild type. Space: indel.

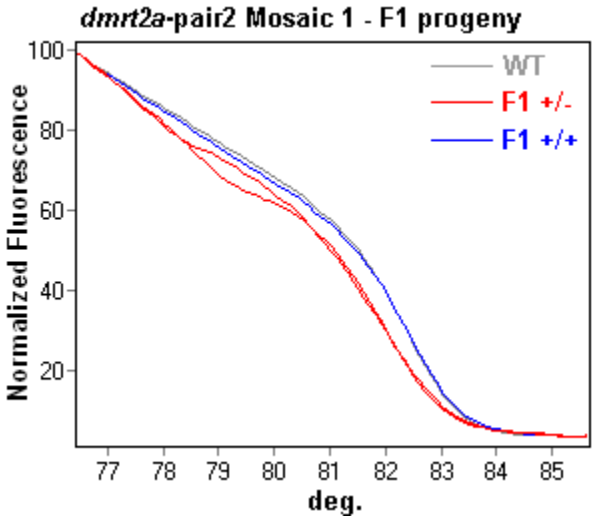


Figure 3.16 High resolution melting analysis (HRMA) of *dmrt2a*-pair2 G0 generation germline transmission. A *dmrt2a*-pair2 mosaic fish was crossed with a wild type one. Individual embryos from F1 progeny were sacrificed so that genomic DNA could be extracted and HRMA could be performed. Samples from embryos not possessing any mutated allele show a melting pattern similar to the wild type (blue). The progeny of *dmrt2a*-pair2 Mosaic 1 transmitted mutant alleles through its germline. It is clear the different between the wild type melting pattern (grey) and the heterozygous ones (red).

```
TCAGGCGGTGTTCACCGGCCCCGGAGAGGACGACACGGGGTCCAAAGACGACGACAAA WT
.......................-----.....................------..... Δ6
.......................--------.....................-------- Δ9
```

Figure 3.17 Sequenced germline transmitted mutations by *dmrt2a*-pair2 G0 generation. Samples that revealed different melting patterns from the wild type (Figure 3.16) were sequenced to confirm the presence of mutant alleles and the extent of each mutation. The binding sites of *dmrt2a*-pair2 are presented in grey. Dot: same nucleotide as the wild type. Space: indel.

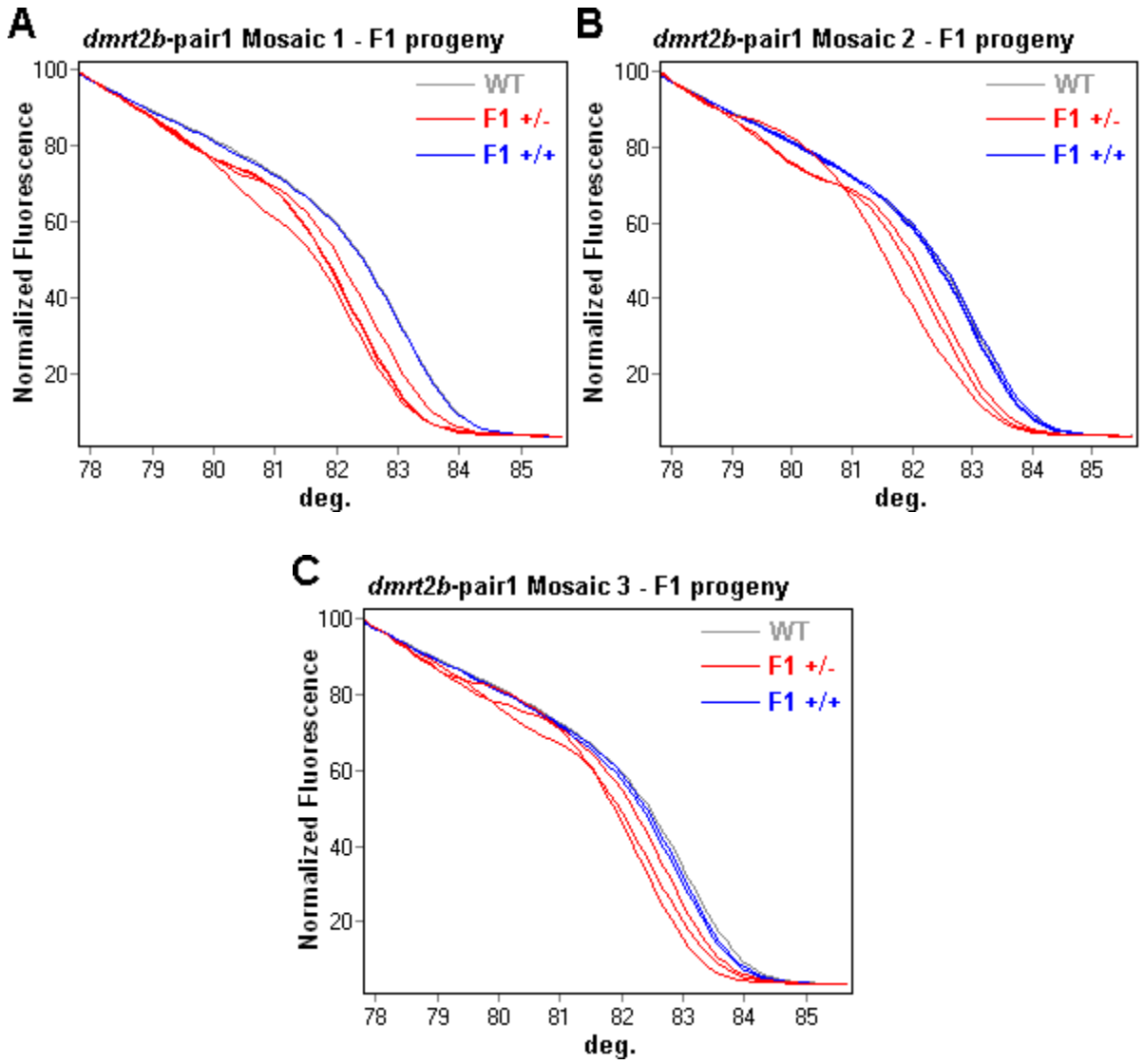
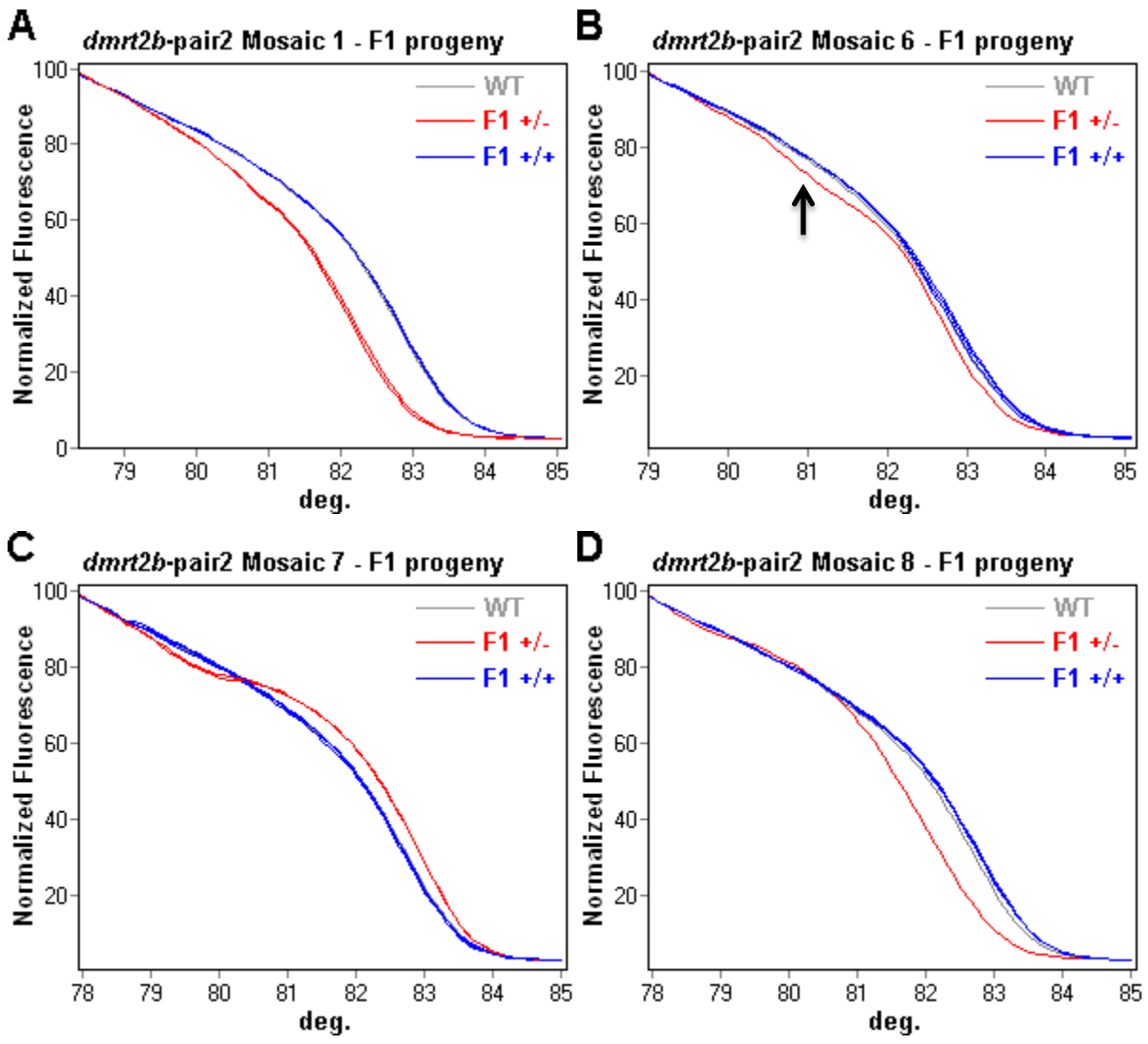


Figure 3.18 High resolution melting analysis (HRMA) of *dmrt2b*-pair1 G0 generation germline transmission. *dmrt2b*-pair1 mosaic fish were crossed with wild type ones. Individual embryos from F1 progeny were sacrificed so that genomic DNA could be extracted and HRMA could be performed. Samples from embryos not possessing any mutant allele show a melting pattern similar to the wild type (blue). All progenies from *dmrt2b*-pair1 mosaic fish transmitted mutant alleles through their germline can be assessed in panels A, B, and C. In these panels, it is clear the different between the wild type melting pattern (grey) and the heterozygous ones (red).

```
TGCGACCCACCTAGAGTGCCTGAAAGCGGAGTGGGTTGCAGACTGCGCCGGACGGGA  WT
.....    .....-----.....    .....                          Δ5
.....    .....------..      .....                          Δ7
.....    .....--.....       .....                          Δ2
.....    ----------------    .....                          Δ19
.....    ...----------------------....                      Δ29
.....    .....------...     .....                          Δ6
```

Figure 3.19 Sequenced germline transmitted mutations by *dmrt2b*-pair1 G0 generation. Samples that revealed different melting patterns from the wild type (Figure 3.18) were sequenced to confirm the presence of mutant alleles and the extent of each mutation. The binding sites of *dmrt2b*-pair1 are presented in grey. Dot: same nucleotide as the wild type. Space: indel.


A
dmrt2b-pair2 Mosaic 1 - F1 progeny
B
dmrt2b-pair2 Mosaic 6 - F1 progeny
C
dmrt2b-pair2 Mosaic 7 - F1 progeny
D
dmrt2b-pair2 Mosaic 8 - F1 progeny
WT
F1 +/-
F1 +/+
Normalized Fluorescence
deg.

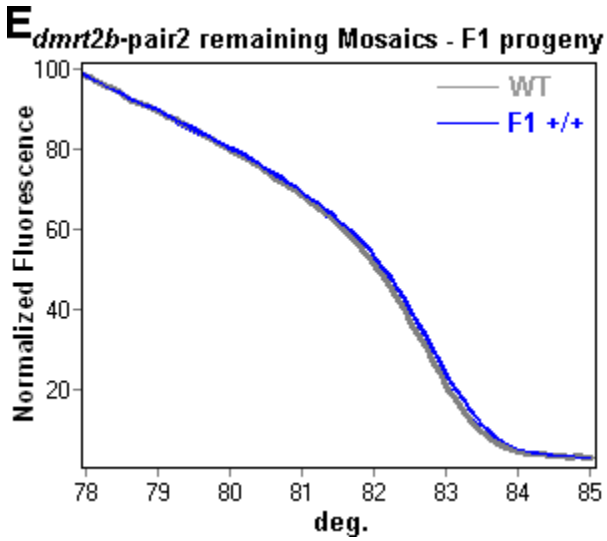


Figure 3.20 High resolution melting analysis (HRMA) of *dmrt2b*-pair2 G0 generation germline transmission. *dmrt2b*-pair2 mosaic fish were crossed with wild type ones. Individual embryos from F1 progeny were sacrificed so that genomic DNA could be extracted and HRMA could be performed. Samples from embryos not possessing any mutated allele show a melting pattern similar to the wild type (blue). The progeny of *dmrt2b*-pair2 mosaic fish that transmit mutant alleles through their germline can be assessed in panels A, B, C, and D. In these panels, it is clear the different between the wild type melting pattern (grey) and the heterozygous ones (red). Some mosaic fish may not transmit mutant alleles through their germline and in this case the melting patterns are similar to the wild type one (panel E). The arrow in panel B points a melting pattern from a genomic DNA sample that only differs from the wild type in 2 bp which shows that even with small deletions, the HRMA is an extremely effective method of genotyping.

```
TGCGACCCACCTAGAGTGCCTGAAAGCGGAGTGGGTTGCAGACTGCGCCGGACGGGA WT
.......................-----.................... Δ5
.......................--...................... Δ2
..GTAA---.....................................  Δ7, +4
```

Figure 3.21 Sequenced germline transmitted mutations by *dmrt2b*-pair2 G0 generation. Samples that revealed different melting patterns from the wild type (Figure 3.20) were sequenced to confirm the presence of mutant alleles and the extent of each mutation. The binding sites of *dmrt2b*-pair2 are presented in grey. Dot: same nucleotide as the wild type. Space: indel.

For all the 4 TALENs pairs injected in G0 embryos, F1 germline mutations could be found. It is important to note that since only a small portion of each F1 progeny was analyzed (in some cases no more than 5 embryos), it would not be accurate to assess the real percentage of G0 adults that transmit mutations through its germline. As described in Figures 3.15, 3.17, 3.19, and 3.21, the lesions induced by single TALEN pairs consist mainly of small indels of 5 to 10 nucleotides with only two exceptions reaching indels 19 and 29 bp for *dmrt2b*-pair1 (Figure 3.19).

Also the utility of HRMA to be used as a more efficient and cost effective genotyping tool was confirmed. Even small lesions as a two bp indel, retrieved a significant shift in the melting pattern as in panel B of Figure 3.20 (arrow). Moreover, wild type melting patterns that were mistaken by heterozygous

ones only happened when the initial concentration prior to the amplification step was significantly different among samples (data not shown).

3.2.4 Homozygous mutants (F2 generation)

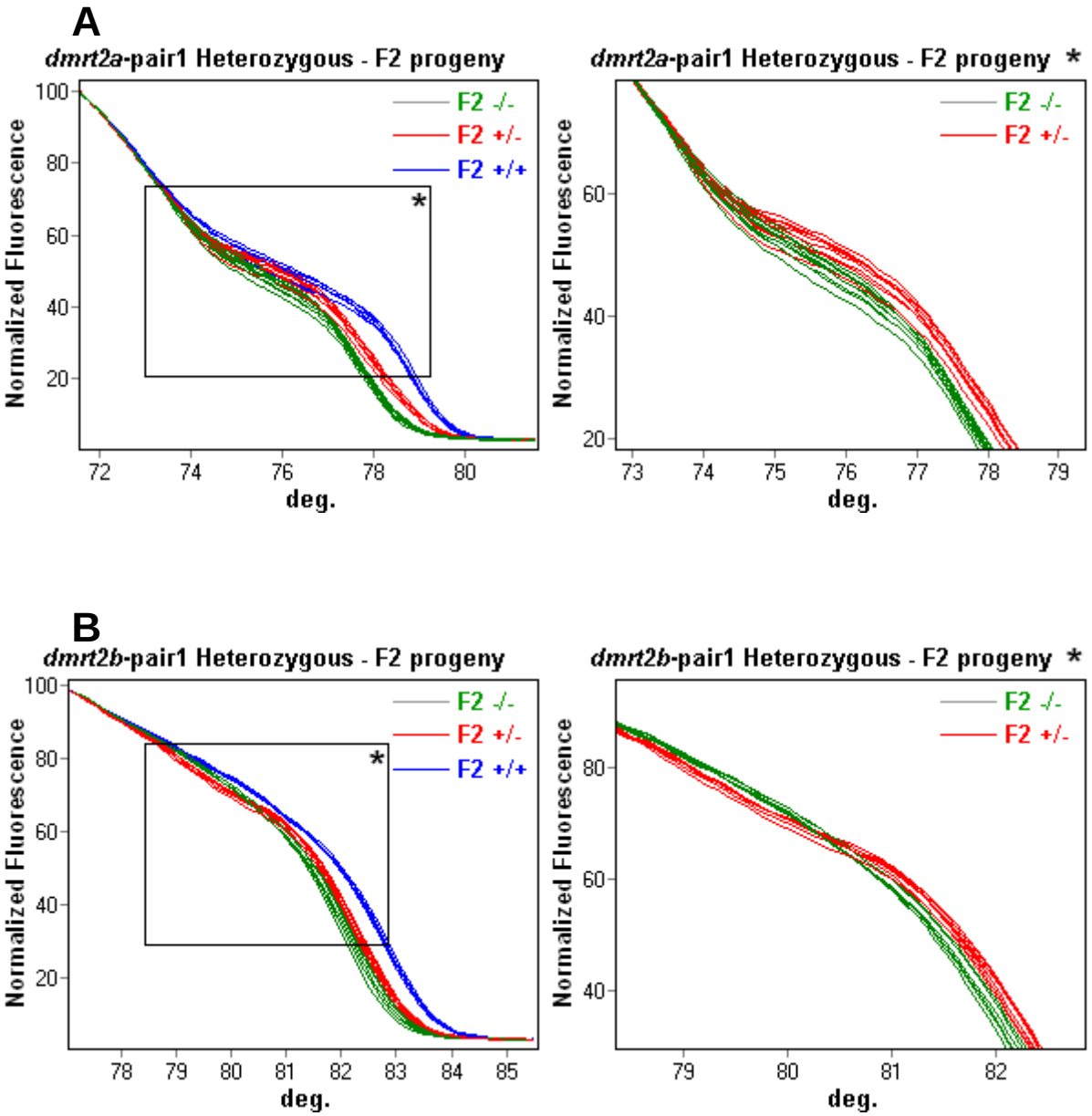
The analysis of mosaics germline transmission revealed several different mutations. Since it would be impossible to raise an F1 generation for each mosaic progeny and develop an F2 generation for each type of mutation found, a choice had to me made, as for which heterozygous zebrafish would be raised. Keeping in mind three conditions, as the ability for the mutation to create a frameshift, the size of the indels generated, and the abundance of each allele within the heterozygous population, it was chosen to raise the *dmrt2a*-pair1 Mosaic 2 progeny (Figure 3.14 panel B) to develop an homozygous mutant for *dmrt2a*, and the *dmrt2b*-pair1 Mosaic 1 and Mosaic 3 progenies (Figure 3.18 panels A and C) to develop an homozygous mutant for *dmrt2b*.

For *dmrt2a* the allele tested so far in homozygoty was one with a 14 bp indel spanning the ATG. Also, this indel creates a frameshift mutation. For *dmrt2b* the allele tested so far in homozygoty was one with a 5 bp indel that creates a frameshift mutation.

3.2.4.1 Failure of homozygous *dmrt2a* and *dmrt2b* mutants to recapitulate its respective MO phenotype

Heterozygous *dmrt2a* and *dmrt2b* fish were incrossed to generate a homozygous mutant for each gene. To assess if the induced deletions would lead to an observable phenotype, heart development was carefully tracked from 32 to 48 hours post fertilization. From published results it was expected that 50% of *dmrt2a* mutant embryos and 46% of *dmrt2b* mutant embryos developed its hearts incorrectly (Saude et al. 2005, Liu et al. 2009, Matsui et al. 2012). Since a heterozygous incross theoretically leads to 25% of mutant embryos, the expected percentages of incorrect heart development for both cases would be of 12.5% and 11.5% for *dmrt2a* and *dmrt2b* respectively. Heart development in both cases resembled a wild type situation with most of the embryos developing their hearts correctly. In a particular experiment, in 82 embryos from *dmrt2a* heterozygous incross, 5 showed incorrect positioning of the heart whereas in 105 embryos from *dmrt2b* heterozygous incross only 2 showed incorrect positioning of the heart. Moreover, the 5 embryos from *dmrt2a* heterozygous incross that developed its heart incorrectly were sacrificed and sequenced, with only 2 possessing the mutant allele.

To test if Mendelian inheritance was being followed, embryos were sacrificed so that its genomic DNA could be extracted and assessed by HRMA (Figure 3.22).



C

	dmrt2a		*dmrt2b*	
	embryos	% of embryos	embryos	% of embryos
Heterozygous	8	33	26	52
Mutant	10*	41	11	22
Wild Type	6	25	13	26

Figure 3.22 High resolution melting analysis (HRMA) of heterozygous progenies reveals the generation of mutant alleles. (A,B) sibling wild type samples are shown in blue. Heterozygous samples are shown in red and mutant samples are shown in green. (A) HRMA of *dmrt2a* heterozygous incross progeny. (B) HRMA of *dmrt2b* heterozygous progeny. (C) Overall percentages of formed alleles.

For both genes the HRMA revealed the consistent generation of mutant alleles which can be seen in panels A and B of Figure 3.22 (In both panels a magnification * of the heterozygous and mutant melting patterns is shown so that the two can be easily distinguished). From panel C of Figure 3.22, the Mendelian inheritance can be easily confirmed for *dmrt2b* heterozygous incross. As for *dmrt2a* the small number of embryos analyzed cannot confirm a Mendelian inheritance. Nevertheless, the number of mutant alleles formed, in this case 10, should theoretically correspond to at least 5 embryos with an incorrect heart development within the experimental batch that was used for this analysis. This did not prove so with only 2 embryos developing its heart incorrectly within the *dmrt2a* heterozygous incross that was used for this analysis.

3.2 Overexpression study of the gene *dmrt2a*

The *dmrt2a* overexpression phenotype has been partially characterized in our lab. Two of its features are the incorrect positioning of the internal organs and the disruption of the synchronized gene expression of the cyclic genes during somitogenesis. In a study where the levels of *dmrt2a* mRNA are indirectly increased by the knockdown of *celf1*, an mRNA binding protein that binds the 3' UTR of *dmrt2a* mRNA promoting its decay, this phenotype has also been described (Matsui et al. 2012). Nevertheless the period during which the cyclic genes are asymmetrically expressed is suggested to cease at the 12-somite stage corroborating with the idea of a time window upon which symmetries and asymmetries of the body plan are defined (Kawakami et al. 2005, Saude et al. 2005).

To evaluate the possibility of maintaining an asymmetric expression of the cyclic genes beyond the 12-somite stage, overexpression of *dmrt2a* was performed by injecting 100pg of *dmrt2a* mRNA per embryo at the one cell stage (approximately three times the amount previously used in our lab). *In situ* hybridization was used to assess the expression of the cyclic genes *deltaC, her7 and her1*.

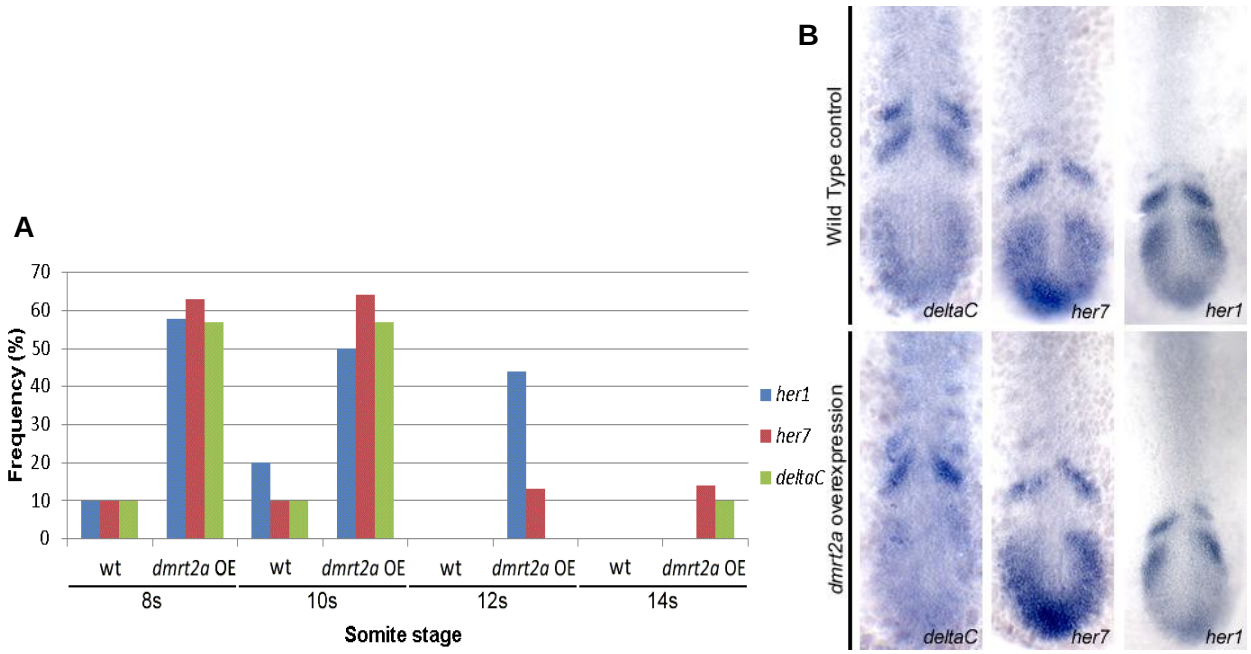


Figure 3.23 The frequency of asymmetric expression of the cyclic genes decreases dramatically after the 12-somite stage. Asymmetric gene expression of the cyclic genes from 8 to 14-somite stage assessed by *in situ* hybridization (A) More than 50% of the analyzed embryos show asymmetric expression of the cyclic genes between 8 to 12-somite stages. This asymmetric gene expression then decreases dramatically with only a small number of embryos showing asymmetric gene expression further on. Although in a lower percentage, some wild type sibling embryos also show asymmetric gene expression before the 12-somite stage. (B) Representative images *of deltaC, her7* and *her1* expression patterns at the 8-somite stage. Upper panels show the wild type sibling controls whereas lower panels show the embryos where *dmrt2a* was overexpressed. Around 450 embryos were used during this experiment.

As can be observed in panel B of Figure 3.23, the overexpression of *dmrt2a* led to an asymmetric expression of the cyclic genes *deltaC*, *her7* and *her1*. Nevertheless, it is perceptible that even when the overexpression of *dmrt2a* is performed with 100pg of *dmrt2a* mRNA per embryo, the cyclic genes regain their symmetric expression after the 12-somite stage (panel A of Figure 3.23). Interestingly, in a wild type situation, asymmetries in the expression of the cyclic genes could also be found prior to the 12-somite stage. These results relate with the existence of a time window after which embryonic asymmetries are already defined.

The work of (Matsui et al. 2012) showed that an increase in the expression of *dmrt2a* leads to failure of the leftward displacement of the heart cone “jogging”. According to (Matsui et al. 2012) wild type zebrafish present abnormal jog in a 2% frequency whereas in a *dmrt2a* increase context this frequency is raised to 28%. This was confirmed by *in situ* hybridization for *cmlc2*, marker of the cardiac myosin light chain (Figure 3.24).

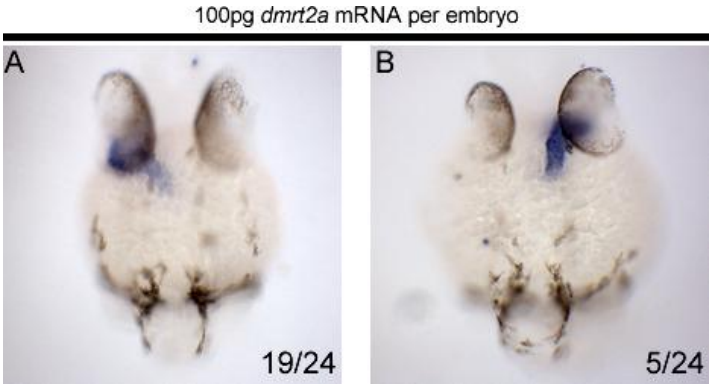


Figure 3.24 Failure of the leftward displacement of the heart cone in a *dmrt2a* overexpression context. (A, B) Representative images of *cmlc2* expression between 28 and 32 hours post fertilization in embryos overexpressing *dmrt2a*. Left "normal" jog was observed in 79% of the studied embryos (A), whereas Right jog could be observed in 21% of the embryos (B).

The embryos where *dmrt2a* was overexpressed sporadically present a kink in notochord at the level of the 12th somite (arrow in Figure 3.25*). Although not yet quantified, it is interesting the appearance of this kink since it coincides with the somite stage where the time window closes and no further asymmetries are defined.

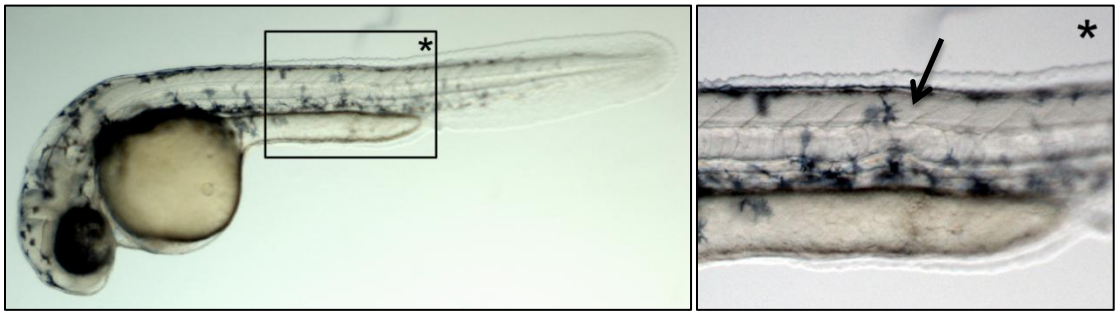


Figure 3.25 Sporadic appearance of a kink in the notochord at the level of the 12th somite. Embryos overexpressing *dmrt2a* sporadically develop a kink in the notochord around the 12th somite.

To evaluate if the failure of achieving prolonged asymmetric expression of the cyclic genes was not due to an insufficient *dmrt2a* overexpression during somite development, *in situ* hybridization was used to compare the expression of *dmrt2a* itself in a wild type situation and in the context of 100pg per embryo mRNA injection, at the onset of the previously referred time window.

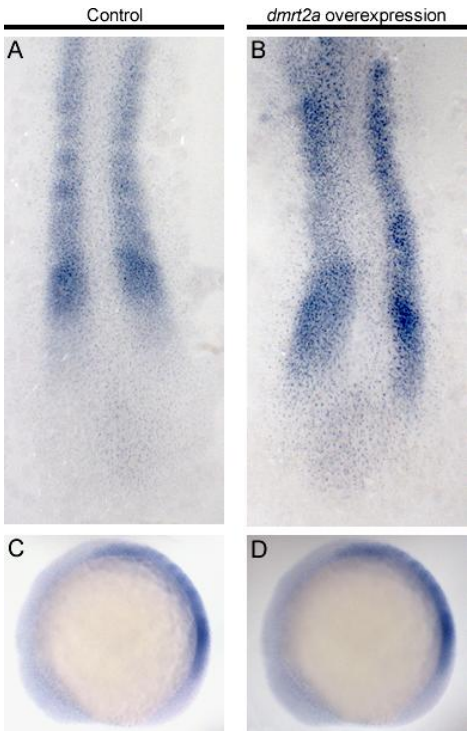


Figure 3.26 Comparison between the expression of *dmrt2a* at the onset of the time window, in a wild type embryo and in an embryo where *dmrt2a* was overexpressed. (A, B, C D) *In situ* hybridization for *dmrt2a* at the 8-somite stage. The expression of *dmrt2a* does not seem to change much between sibling wild type controls (A, C) and the *dmrt2a* overexpression embryos (B, D). (A, B) Flat mount view at the level of the somites. (C, D) Whole mount dorsal view.

Although a significant change in the development of the somites in the context of *dmrt2a* overexpression is seen in panel B of Figure 3.26, the overall expression of *dmrt2a* does not seem to change as much as it would be expected between the two situations.

Therefore, it was decided to increase the amount of *dmrt2a* mRNA per embryo to 200pg. The leftward displacement of the heart cone was again assessed as previously described.

When one-cell stage embryos were injected with a significant higher amount of *dmrt2a* mRNA a complete randomization of the heart cone displacement could be observed (Figure 3.27). Also, in two independent experiments, two thirds of the embryos injected with this amount of *dmrt2a* mRNA, develop a curved body either to the left or to the right (Figure 3.28).

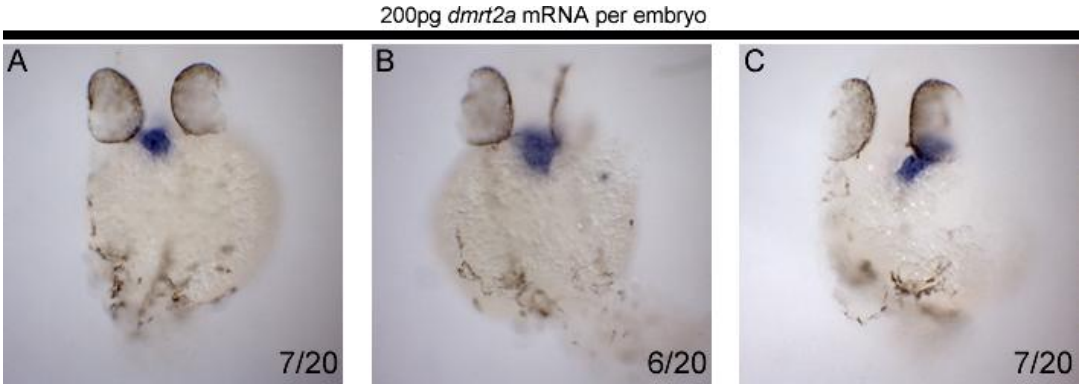


Figure 3.27 Complete randomization of the heart cone displacement. (A, B) Representative images of *cmlc2* expression between 28 and 32 hours post fertilization in embryos where *dmrt2a* was overexpressed by the injection of 200pg *dmrt2a* mRNA at the one cell stage. Left "normal" jog was observed in 35% of the studied embryos (A), no jog could be observed in 30% of the embryos (B) and right jog was observed in 35% embryos (C).

Figure 3.28 Zebrafish larvae present a curved body either to the left or to the right. 5 days post fertilization zebrafish larvae that were previously injected with 200pg of *dmrt2a* mRNA show a curved body. This curvature is completely randomized with 36% of the larvae showing no curvature (A), 31% showing a curvature to the left side (B), and 33% showing a curvature to the right (C). Two independent experiments were performed.

Although not yet assessed, it is of the most interest to investigate if the stronger phenotype obtained with a higher level of *dmrt2a* overexpression correlates with an expansion of the time window. Possibly, a delay in the time window closure, leads to an asymmetrical formation of several somites. This would eventually lead to the development of a curved body, as seen in (Figure 3.28).

Chapter 4

Discussion and Future work

4.1 TALENs, as a new tool to generate mutant alleles in zebrafish

The TALEN technology was used to develop zebrafish mutant alleles for the genes *dmrt2a* and *dmrt2b.* The ability of TALENs to introduce genomic deletions should be highlighted with the two key features attributed to TALENs, efficiency and specificity (Cermak et al. 2011, Cade et al. 2012, Dahlem et al. 2012), being strongly confirmed in the course of this work.

Considering that each TALEN pair is composed of two independent monomers that have to bind and cleave DNA, it is amazing how all the eight monomers designed and assembled during this work achieved this task. Also, table 3.1 clearly shows that depending on the amount of TALENs mRNA injected, the percentage of fish possessing mosaic mutations could reach more than 90%. All injected TALEN pairs successfully introduced mosaic mutations into the zebrafish genome, consisting mainly on small indels of 5 to 10 bp, with two exceptions being of 19 and 29 bp indel introduced by the same TALEN pair. These mutations were introduced specifically at the predicted target site and even when higher amounts of TALENs mRNA were used, this specificity was not lost.

The ability of multiple TALEN pairs to be used together in order to achieve bigger genomic deletions was recently published (Xiao et al. 2013). In 3.1.3.2 this was assessed with the co-injection of four TALEN pairs, comprising two distinct target sites distancing 100 bp from each other. The successful amplification through PCR of small DNA fragments that lacked these 100 bp confirmed it. Furthermore, the approach that was used to assess this (described in 3.1.3.2), without a cloning step to separate each DNA fragment, provides an easier and quicker way to confirm multiple TALEN activity when big genomic deletions are intended. During the process of G0 Mosaic generation, the exact notion of each created mutation, which can be obtained by a cloning step and subsequent sequencing, is not that relevant, since the information on which of these mutations will pass through the germline is not known. That way, the simple notion that a big DNA deletion was obtained in a particular genomic region can be sufficient to confirm the activity of the injected TALENs.

Another feature that had already been attributed to TALENs is their low toxicity to zebrafish embryonic cells (Cade et al. 2012). Although not quantified during this work this could be observed since no significant death seemed to occur even when 400pg of TALENs mRNA were injected as in the cases of *dmrt2b*-pair2 (3.1.3.1) and *dmrt2a*-pair1+pair2 (3.1.3.2).

It is also important to remember that TALENs design and assembly revealed to be a simple process. TALENs could be easily designed when following NICHD Zebrafish Core rules and although the assembling process lasted longer than the 5 days reported by (Cermak et al. 2011) (approximately 3 weeks for all four TALEN pairs), most of the cloning reactions worked at the first attempt.

Taken together these results confirmed TALENs as ideal tools to introduce genomic deletions into the zebrafish genome.

4.2 Zebrafish TALEN mutants phenotype *versus* zebrafish morphants phenotype

Two TALEN zebrafish mutant alleles were produced for the genes *dmrt2a* and *dmrt2b*. Both mutations introduced a frame shift in the open reading frame of these genes. More specifically the *dmrt2a* mutant has a 14 bp indel spanning the annotated ATG start codon, whereas the *dmrt2b* mutant has a 5 bp indel 60 bp downstream of the annotated ATG start codon. Nevertheless, no heart positioning phenotype was observed that could resemble the published ones using the MO technology (Saude et al. 2005, Liu et al. 2009, Matsui et al. 2012).

Before any considerations are made, it should be noted, that the protein levels of Dmrt2a and Dmrt2b have not yet been assessed in these mutants. Depending on the ability of these mutants to produce functional proteins distinct possibilities arise.

If it proves that these mutants are lacking Dmrt2a and Dmrt2b proteins respectively, it has to be assessed if any kind of compensation mechanism is acting in a way that the expected phenotype is absent, or else, if the described MO phenotypes consist mainly on artifacts. On the other side, if these mutants are still able to produce a functional protein then a question arises: how is this possible since a frame shift is introduced in both sequences?

From (Matsui et al. 2012), it seems that the *dmrt2a* Morphant phenotype is consistent. In this work the levels of *dmrt2a* are indirectly manipulated through the overexpression of an RNA binding protein that binds the 3' UTR of *dmrt2a* mRNA promoting its degradation. Taken that the phenotype observed is the same as in (Saude et al. 2005), then the possibility that the *dmrt2a* Morphant phenotype is somehow an artifact is extremely reduced. In the case of *dmrt2b* morphants, its phenotype was only assessed in (Liu et al. 2009). In this study, the rescue experiment with *dmrt2b* mRNA was not fully accomplished, with only 50% of the studied embryos developing normally. It has been shown that MOs can have off Target effects mediated by the Tp53 pathway, and not exclusively related to cell death which may lead to phenotype misinterpretation (Gerety and Wilkinson 2011). Nevertheless, the Morphant study of *dmrt2b* gene uses a p53-control MO with no phenotype being observed in this case (Liu et al. 2009).

The study of the *dmrt2b* morphants also addresses the question of if these two genes can compensate for each other. In this study neither *dmrt2a* mRNA nor *dmrt2b* mRNA can rescue each other's morphants phenotypes. During this work in situ hybridization was further used to analyze the expression of *dmrt2a* in the *dmrt2b* mutants but no change in the expression levels was observed. Restricted to the case of the *dmrt2a* mutants, it seems unlikely that a compensation mechanism is happening. In (Matsui et al. 2012), the mRNA levels of *dmr2a* mRNA are assessed through RT-PCR. When *dmrt2a* is indirectly decreased by the overexpression of *celf1*, as already explained, a phenotype similar to the *dmrt2a* Morphant can be described. The decrease in the *dmrt2a* mRNA levels that lead to this phenotype correspond to 50% of the *dmrt2a* mRNA levels assessed in a wild type control. This would mean that only

below this level, the compensation mechanisms would be triggered which seems unlikely. Left-right establishment has been conserved throughout evolution, and so it should be tightly controlled.

It has been reported by the NICHD Zebrafish Core (NICHD) that several Labs that have been developing TALEN zebrafish mutants have been failing to recapitulate previously described Morphant phenotypes and that most of the successful cases arise from big genomic lesions from 100 to thousands of bp. This would clearly point in the direction that maybe zebrafish could be using alternative forms of splicing or alternative ATGs (Jezewski et al. 2009), or even stop codon readthrough (Williams et al. 2004, Jungreis et al. 2011) in order to produce the desired proteins (Xiao et al. 2013). This idea would explain the ability of our *dmrt2a* mutant to develop a functional protein since this mutant has an ATG in frame at the 5' UTR. Also, a stop codon had to be readthrough in order to produce the functional protein. It is important to note that these assumptions are made on the fact that the integrity of the DM domain of Dmrt2a has to be kept, since it has been published that the Nuclear localization signal of this protein resides in the DM domain (Zhang et al. 2001). In the case of the *dmrt2b* mutant developed during this work, it's difficult to accept this previous idea since the mutation introduced is located 56 bp downstream of the annotated ATG which would let the ribosome to start translation normally. Still, downstream of the mutation and prior to the start of the DM domain, two ATGs could be used to produce a functional protein, once again, reading through stop codons. For both cases, an analysis of the Expressed Sequence Tags was done, but this did not reveal possible alternative transcripts that could fit in the studied cases.

Although only these two zebrafish mutants have yet been developed, this work left the possibility for other *dmrt2a* and *dmrt2b* mutant alleles to be tested. As for *dmrt2a*, fish developed with *dmrt2a*-pair2, which introduces in the *dmrt2a* gene lesions similar to the one tested for *dmrt2b* are already in a F1 generation and will be soon incrossed to generate homozygous mutants for this mutations. Also, in section 3.1.3.2., big genomic lesions of 100 bp were introduced in the *dmrt2a* gene. Fish possessing these lesions are still in its G0 generation but will be used as well to generate homozygous mutants. In what concerns to dmrt2b, it's referred in section 3.2.4, that heterozygous fish that possess 19 and 29 bp indels are being used to generate homozygous mutants for these mutations. It will be of great interest to know whether any of these forthcoming *dmrt2a* and *dmrt2b* mutants will recapitulate the morphants phenotypes and possibly give insights on what kind of mechanisms zebrafish is using to deal with the so far introduced mutations. If it proves that only big genomic lesions can successfully knock out these genes, one extra *dmrt2b* TALEN pair with a target site distant from the ones targeted in this work for this gene can be designed and assembled, so that the experiment from 3.1.3.2., can be repeated for *dmrt2b*.

Upon successful development and characterization of homozygous *dmrt2a* and *dmrt2b* mutants, the generation of a double *dmrt2a dmrt2b* mutant will help us further study the evolution and divergence of these two paralogous genes and to what extent are they related with the homologous mouse *dmrt2*.

Finally it should be remember that many known mutations in the zebrafish genome arouse from ENU-based forward genetics screens where random mutations were introduced into the zebrafish genome with the subsequent progenies being screen in terms of phenotype (Huang et al. 2012). It is possible that many of these random mutations, although introducing a frameshift, did not led to an

observable phenotype which could explain the problems we are facing now, not understanding how zebrafish copes with mutations that introduce a frame shift. The use of new genetic tools to manipulate the zebrafish genome in a reverse genetics approach will help to clarify the mechanisms behind this question.

4.3 An overexpression analysis reveals a time window of action of *dmrt2a*

It has been suggested that a time window of opportunity during which asymmetries are defined exists during 6 to 13-somite stage in zebrafish (Kawakami et al. 2005, Saude et al. 2005). To test this in the context of *dmrt2a* overexpression, *dmrt2a* mRNA was injected at the one cell stage and cyclic gene expression during 8 to 14-somite stage was assessed. Asymmetric gene expression of these genes, was observed clearly between 8 and 12-somite stages, but found to become symmetric afterwards. Nevertheless, and despite regaining a symmetric gene expression, a percentage of embryos higher than the previously published wild type situation (Matsui et al. 2012), failed to correctly displace its heart cone to the left. This reinforces the idea that a time window during which asymmetries are defined do exists in this period.

Interestingly, embryos overexpressing *dmrt2a*, sporadically developed a kink in the notochord around somite 12. This has not yet been quantified, but may be related with the fact that during this period, zebrafish somites develop asymmetrically between both sides of the left-right axis. It is possible that a mechanism of reestablishment of symmetry during this period, repositions these somites along with the notochord leading to the observed kink.

Although the results indicated the existence of a time window, still, it could be that the levels of mRNA used to overexpress *dmrt2a* were simply not sufficient to prologue the asymmetric gene expression of the cyclic genes further that the 12-somite stage. Strikingly, when in situ hybridization was used to assess the levels of *dmrt2a* expression between the experimental and control situations at the onset of this time window these did not reveal a significant change. To further test this hypothesis, a significant higher amount of mRNA was used to overexpress *dmrt2a* and although the cyclic gene expression during the period of the time window has not yet been assessed, the effects of this higher overexpression could be observed by the complete randomization of the heart cone displacement. Moreover, two thirds of the embryos exposed to a significant higher amount of *dmrt2a* mRNA, developed a curved body either to the left or right sides. It would be of great interest to analyze if the somite region where these embryos start curving their body corresponds to the same region where a kink in the notochord was observed. If it corresponds, and if it is observed a prolonged time window in this case, then possibly the reason is that these embryos fail to reposition the asymmetrically formed somites which inevitably would lead to the observed curved body.

Although these experiments retrieved interesting data, it should be noted that these are ongoing experiments. The appropriate controls should be used and the analysis and quantifications of the

observed kink in the notochord have to be made. Moreover, the study on the time window has to be carefully repeated for this last situation where a higher level of overexpression was used.

Chapter 5

Conclusion

During this work TALEN technology was successfully used to generate mutant alleles for *dmrt2a* and *dmrt2b.* The efficiency and specificity, as well as the ability of this technology to generate genomic deletions from 2 bp up to more than 100 bp were strongly confirmed.

Although the mutated alleles tested so far in homozygoty did not induce any observable phenotype, this work laid bases for several different alleles to be tested in the future.

Nevertheless, the results so far, raised interesting questions as to what extent are the studied MO phenotypes accurate or else what molecular mechanism(s) is zebrafish using to cope with frameshift mutations.

The possible existence of a time window of opportunity during which asymmetries are defined was also assessed and confirmed during this work. Future experiments will confirm if and how is *dmrt2a* involved in this process.

Bibliography

Babu, D. and S. Roy (2013). "Left-right asymmetry: cilia stir up new surprises in the node." Open Biol 3(5): 130052.

Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt and U. Bonas (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." Science 326(5959): 1509-1512.

Boettger, T., L. Wittler and M. Kessel (1999). "FGF8 functions in the specification of the right body side of the chick." Curr Biol 9(5): 277-280.

Bogdanove, A. J., S. Schornack and T. Lahaye (2010). "TAL effectors: finding plant genes for disease and defense." Curr Opin Plant Biol 13(4): 394-401.

Bradley, K. M., J. B. Elmore, J. P. Breyer, B. L. Yaspan, J. R. Jessen, E. W. Knapik and J. R. Smith (2007). "A major zebrafish polymorphism resource for genetic mapping." Genome Biol 8(4): R55.

Cade, L., D. Reyon, W. Y. Hwang, S. Q. Tsai, S. Patel, C. Khayter, J. K. Joung, J. D. Sander, R. T. Peterson and J. R. Yeh (2012). "Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs." Nucleic Acids Res 40(16): 8001-8010.

Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove and D. F. Voytas (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." Nucleic Acids Res 39(12): e82.

Dahlem, T. J., K. Hoshijima, M. J. Jurynec, D. Gunther, C. G. Starker, A. S. Locke, A. M. Weis, D. F. Voytas and D. J. Grunwald (2012). "Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome." PLoS Genet 8(8): e1002861.

Dequeant, M. L. and O. Pourquie (2008). "Segmental patterning of the vertebrate embryonic axis." Nat Rev Genet 9(5): 370-382.

Doyle, E. L., N. J. Booher, D. S. Standage, D. F. Voytas, V. P. Brendel, J. K. Vandyk and A. J. Bogdanove (2012). "TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction." Nucleic Acids Res 40(Web Server issue): W117-122.

Engler, C., R. Gruetzner, R. Kandzia and S. Marillonnet (2009). "Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes." PLoS One 4(5): e5553.

Fine, E. Retrieved 17-11-2013, 2013, from http://baolab.bme.gatech.edu/Research/BioinformaticTools/assembleTALSequences.html.

Fliegauf, M., T. Benzing and H. Omran (2007). "When cilia go bad: cilia defects and ciliopathies." Nat Rev Mol Cell Biol 8(11): 880-893.

Gerety, S. S. and D. G. Wilkinson (2011). "Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development." Dev Biol 350(2): 279-289.

Holley, S. A., D. Julich, G. J. Rauch, R. Geisler and C. Nusslein-Volhard (2002). "her1 and the notch pathway function within the oscillator mechanism that regulates zebrafish somitogenesis." Development 129(5): 1175-1183.

Hollway, G. E., R. J. Bryson-Richardson, S. Berger, N. J. Cole, T. E. Hall and P. D. Currie (2007). "Whole-somite rotation generates muscle progenitor cell compartments in the developing zebrafish embryo." Dev Cell 12(2): 207-219.

Hong, C. S., B. Y. Park and J. P. Saint-Jeannet (2007). "The function of Dmrt genes in vertebrate development: it is not just about sex." Dev Biol 310(1): 1-9.

Huang, P., Z. Zhu, S. Lin and B. Zhang (2012). "Reverse genetic approaches in zebrafish." J Genet Genomics 39(9): 421-433.

Integrated DNA Technologies, I. Retrieved 17-11-2013, 2013, from http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/.

Ishimatsu, K., K. Horikawa and H. Takeda (2007). "Coupling cellular oscillators: a mechanism that maintains synchrony against developmental noise in the segmentation clock." Dev Dyn 236(6): 1416-1421.

Jezewski, P. A., P. K. Fang, T. L. Payne-Ferreira and P. C. Yelick (2009). "Alternative splicing, phylogenetic analysis, and craniofacial expression of zebrafish tbx22." Dev Dyn 238(6): 1605-1612.

Jungreis, I., M. F. Lin, R. Spokony, C. S. Chan, N. Negre, A. Victorsen, K. P. White and M. Kellis (2011). "Evidence of abundant stop codon readthrough in Drosophila and other metazoa." Genome Res 21(12): 2096-2113.

Kawakami, K. (2005). "Transposon tools and methods in zebrafish." Dev Dyn 234(2): 244-254.

Kawakami, Y., A. Raya, R. M. Raya, C. Rodriguez-Esteban and J. C. Izpisua Belmonte (2005). "Retinoic acid signalling links left-right asymmetric patterning and bilaterally symmetric somitogenesis in the zebrafish embryo." Nature 435(7039): 165-171.

Kishimoto, N., Y. Cao, A. Park and Z. Sun (2008). "Cystic kidney gene seahorse regulates cilia-mediated processes and Wnt pathways." Dev Cell 14(6): 954-961.

Levin, M. (2004). "The embryonic origins of left-right asymmetry." Crit Rev Oral Biol Med 15(4): 197-206.

Lewis, J., A. Hanisch and M. Holder (2009). "Notch signaling, the segmentation clock, and the patterning of vertebrate somites." J Biol 8(4): 44.

Liu, S., Z. Li and J. F. Gui (2009). "Fish-specific duplicated dmrt2b contributes to a divergent function through Hedgehog pathway and maintains left-right asymmetry establishment function." PLoS One 4(9): e7261.

Lourenco, R., S. S. Lopes and L. Saude (2010). "Left-right function of dmrt2 genes is not conserved between zebrafish and mouse." PLoS One 5(12): e14438.

Lourenço, R. and L. Saúde (2010). "Symmetry OUT, Asymmetry IN." Symmetry 2(2): 1033-1054.

Matsui, T., A. Sasaki, N. Akazawa, H. Otani and Y. Bessho (2012). "Celf1 regulation of dmrt2a is required for somite symmetry and left-right patterning during zebrafish development." Development 139(19): 3553-3560.

McGrath, J., S. Somlo, S. Makova, X. Tian and M. Brueckner (2003). "Two populations of node monocilia initiate left-right asymmetry in the mouse." Cell 114(1): 61-73.

Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L.

Zhang, P. D. Gregory and E. J. Rebar (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol 29(2): 143-148.

Morillas, H. N., M. Zariwala and M. R. Knowles (2007). "Genetic causes of bronchiectasis: primary ciliary dyskinesia." Respiration 74(3): 252-263.

Murphy, M. W., D. Zarkower and V. J. Bardwell (2007). "Vertebrate DM domain proteins bind similar DNA sequences and can heterodimerize on DNA." BMC Mol Biol 8: 58.

Nakamura, T. and H. Hamada (2012). "Left-right patterning: conserved and divergent mechanisms." Development 139(18): 3257-3262.

NCBI. Retrieved 17-11-2013, 2013, from http://www.ncbi.nlm.nih.gov/tools/primer-blast/.

NICHD. "NICHD." Retrieved 17-11-13, 2013, from https://science.nichd.nih.gov/confluence/display/zcore/Home++Welcome+to+the+NICHD+Zebrafish+Core.

Nonaka, S., H. Shiratori, Y. Saijoh and H. Hamada (2002). "Determination of left-right patterning of the mouse embryo by artificial nodal flow." Nature 418(6893): 96-99.

Nonaka, S., Y. Tanaka, Y. Okada, S. Takeda, A. Harada, Y. Kanai, M. Kido and N. Hirokawa (1998). "Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein." Cell 95(6): 829-837.

Palmeirim, I., D. Henrique, D. Ish-Horowicz and O. Pourquie (1997). "Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis." Cell 91(5): 639-648.

Pornprasert, S., A. Phusua, S. Suanta, R. Saetung and T. Sanguansermsri (2008). "Detection of alpha-thalassemia-1 Southeast Asian type using real-time gap-PCR with SYBR Green1 and high resolution melting analysis." Eur J Haematol 80(6): 510-514.

Price, E. P., H. Smith, F. Huygens and P. M. Giffard (2007). "High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of Campylobacter jejuni." Appl Environ Microbiol 73(10): 3431-3436.

Riedel-Kruse, I. H., C. Muller and A. C. Oates (2007). "Synchrony dynamics during initiation, failure, and rescue of the segmentation clock." Science 317(5846): 1911-1915.

Rodriguez-Esteban, C., J. Capdevila, Y. Kawakami and J. C. Izpisua Belmonte (2001). "Wnt signaling and PKA control Nodal expression and left-right determination in the chick embryo." Development 128(16): 3189-3195.

Sander, J. D., L. Cade, C. Khayter, D. Reyon, R. T. Peterson, J. K. Joung and J. R. Yeh (2011). "Targeted gene disruption in somatic zebrafish cells using engineered TALENs." Nat Biotechnol 29(8): 697-698.

Saude, L., R. Lourenco, A. Goncalves and I. Palmeirim (2005). "terra is a left-right asymmetry gene required for left-right synchronization of the segmentation clock." Nat Cell Biol 7(9): 918-920.

Segalen, M., C. A. Johnston, C. A. Martin, J. G. Dumortier, K. E. Prehoda, N. B. David, C. Q. Doe and Y. Bellaiche (2010). "The Fz-Dsh planar cell polarity pathway induces oriented cell division via Mud/NuMA in Drosophila and zebrafish." Dev Cell 19(5): 740-752.

Seo, K. W., Y. Wang, H. Kokubo, J. R. Kettlewell, D. A. Zarkower and R. L. Johnson (2006). "Targeted disruption of the DM domain containing transcription factor Dmrt2 reveals an essential role in somite patterning." Dev Biol 290(1): 200-210.

Sobek, J., K. Bartscherer, A. Jacob, J. D. Hoheisel and P. Angenendt (2006). "Microarray technology as a universal tool for high-throughput analysis of biological systems." Comb Chem High Throughput Screen 9(5): 365-380.

Speder, P., A. Petzoldt, M. Suzanne and S. Noselli (2007). "Strategies to establish left/right asymmetry in vertebrates and invertebrates." Curr Opin Genet Dev 17(4): 351-358.

Stickney, H. L., M. J. Barresi and S. H. Devoto (2000). "Somite development in zebrafish." Dev Dyn 219(3): 287-303.

Valente, E. M., C. V. Logan, S. Mougou-Zerelli, J. H. Lee, J. L. Silhavy, F. Brancati, M. Iannicelli, L. Travaglini, S. Romani, B. Illi, M. Adams, K. Szymanska, A. Mazzotta, J. E. Lee, J. C. Tolentino, D. Swistun, C. D. Salpietro, C. Fede, S. Gabriel, C. Russ, K. Cibulskis, C. Sougnez, F. Hildebrandt, E. A. Otto, S. Held, B. H. Diplas, E. E. Davis, M. Mikula, C. M. Strom, B. Ben-Zeev, D. Lev, T. L. Sagie, M. Michelson, Y. Yaron, A. Krause, E. Boltshauser, N. Elkhartoufi, J. Roume, S. Shalev, A. Munnich, S. Saunier, C. Inglehearn, A. Saad, A. Alkindy, S. Thomas, M. Vekemans, B. Dallapiccola, N. Katsanis, C. A. Johnson, T. Attie-Bitach and J. G. Gleeson (2010). "Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes." Nat Genet 42(7): 619-625.

Vandenberg, L. N. and M. Levin (2010). "Far from solved: a perspective on what we know about early mechanisms of left-right asymmetry." Dev Dyn 239(12): 3131-3146.

Vandenberg, L. N. and M. Levin (2013). "A unified model for left-right asymmetry? Comparison and synthesis of molecular models of embryonic laterality." Dev Biol 379(1): 1-15.

Vermot, J., J. Gallego Llamas, V. Fraulob, K. Niederreither, P. Chambon and P. Dolle (2005). "Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo." Science 308(5721): 563-566.

Vermot, J. and O. Pourquie (2005). "Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos." Nature 435(7039): 215-220.

Vilhais-Neto, G. C., M. Maruhashi, K. T. Smith, M. Vasseur-Cognet, A. S. Peterson, J. L. Workman and O. Pourquie (2010). "Rere controls retinoic acid signalling and somite bilateral symmetry." Nature 463(7283): 953-957.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio).

Williams, I., J. Richardson, A. Starkey and I. Stansfield (2004). "Genome-wide prediction of stop codon readthrough during translation in the yeast Saccharomyces cerevisiae." Nucleic Acids Res 32(22): 6605-6616.

Xiao, A., Z. Wang, Y. Hu, Y. Wu, Z. Luo, Z. Yang, Y. Zu, W. Li, P. Huang, X. Tong, Z. Zhu, S. Lin and B. Zhang (2013). "Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish." Nucleic Acids Res 41(14): e141.

Zhang, L., Z. Hua, J. Ren and A. Meng (2001). "The nuclear localization signal of zebrafish terra is located within the DM domain." FEBS Lett 503(1): 25-29.